Assessment of a measles outbreak response vaccination campaign, and two measles parameter estimation methods

by

James Mba Azam

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Mathematics in the Faculty of Science at Stellenbosch University

Department of Mathematical Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisor: Prof. Juliet R.C. Pulliam

March 2018

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: .....
James Mba Azam

Date: March 2018 .....

Abstract

Assessment of a measles outbreak response vaccination campaign, and two measles parameter estimation methods

James Mba Azam

Department of Mathematical Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MSc. (Mathematics)

March 2018

Measles is highly transmissible, and is a leading cause of vaccine-preventable death among children. Consequently, it is regarded as a public health issue worldwide and has been targeted for elimination by 5 out of the 6 WHO regions by 2020, the exception being the WHO Africa region. The hope of achieving this target, however, seems bleak as regular outbreaks continue to occur. Data from these outbreaks are useful for pursuing important questions about measles dynamics and control. This thesis is structured to investigate two questions: the first is on how well the time series susceptible-infected-recovered (TSIR) model and removal method perform when they are used to estimate parameters from poor quality data on measles epidemics. We simulate stochastic epidemics for four spatial patches, resembling data that are collected in low-income countries where resources are limited for properly collecting and reporting data on measles epidemics. We then obtain from the simulated data sets, the size of the initial susceptible population S_0, and the basic reproductive ratio $\mathcal{R}_0$ - for the TSIR; and S_0, and either the effective reproduction number $\mathcal{R}_{eff}$, or the basic reproductive ratio $\mathcal{R}_0$ - for the removal method, depending on the simulation assumptions. To assess performance, we quantify the biases that result when we tweak some of the simulation assumptions and modify the data to ensure it is in a form usable for each of the two methods. We find that

the performance of the methods depends on the assumptions underlying the data generation process, the degree of spatial aggregation, the chosen method of modifying the data to put it in a form usable for the estimation method, and the parameter being fitted. The removal method S_0 estimates at the patch level are almost unbiased when the population is naive, but are biased when aggregated to the population level, whether the population is initially naive or not. Furthermore, the removal $\mathcal{R}_0$ and $\mathcal{R}_{eff}$ estimates are generally biased. The TSIR model, on the other hand, seems more robust in estimating both S_0 and $\mathcal{R}_0$ for non-naive populations. These findings are useful because they give us an idea of the biases in the fits of these methods to actual data of the same nature as the simulated epidemics.

For the second question, we assess the impact of an outbreak response vaccination campaign which was organised in reaction to a measles outbreak in an all-boys high school in Stellenbosch, South Africa. We achieve this by formulating a discrete stochastic susceptible-exposed-infected-recovered (SEIR) model with daily time-steps, ignoring births and deaths. Using the model, we analyse multiple scenarios that allow us to estimate the cases averted, and to predict the cases remaining until the epidemic ended, and the time frame within which those cases would occur. Summarizing across scenarios, we estimate that a median of 255 cases (range $60 - 493$) were averted. Also, a median of 15 remaining cases (range $1 - 33$), and a median of 4 remaining weeks (range $1 - 16$) were expected until the epidemic ended. We conclude that the campaign was successful in averting many potential cases.

Keywords: measles, outbreak response vaccination, time series susceptible-infected-recovered model, removal method, parameter estimation.

Opsomming

Evaluering van 'n inentings veldtog na 'n masels uitbraak, en twee metodes om masels parameters te beraam

James Mba Azam

Departement Wiskundige Wetenskappe,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Tesis: MSc. (Wiskunde)

Maart 2018

Masels is hoogs oordraagbaar, en is 'n leidende oorsaak van entstof-voorkombare sterftes onder kinders. Gevolglik word dit beskou as 'n wêreldwy openbare gesondheidskwessie en word daar beoog om teen 2020 die virus in 5 uit die 6 WHO-streke te elimineer, met uitsondering die WGO-Afrika-streek. Die kans om hierdie teiken te bereik, lyk egter skraal aangesien gereelde uitbrake steeds voorkom. Data van hierdie uitbrake is nuttig om belangrike vrae oor masels-dinamika en beheer te ondersoek. Hierdie proefskrif ondersoek twee vrae. Die eerste is hoe goed die tydreeks vatbaar-aansteeklik-herstel (TSIR) model en die verwyderings metode presteer wanneer dit gebruik word om parameters met behulp van data van swak gehalte oor maselsepidemies te skat. Ons simuleer stogastiese epidemies vir vier areas, wat ooreenstem met data wat in lae-inkomste lande versamel word, waar hulpbronne vir die behoorlike versameling en rapportering van data oor maselsepidemies beperk is. Ons kry dan uit die gesimuleerde datastelle, die grootte van die aanvanklike vatbare populasie S_0, en die basiese reproduktiewe verhouding $\mathcal{R}_0$ - vir die TSIR; en S_0, en óf die effektiewe reproduktiewe getal $\mathcal{R}_{eff}$, of die basiese reproduktiewe verhouding $\mathcal{R}_0$ - vir die verwydering metode, afhangende van die simulasie aannames. Om prestasie te evalueer, bereken ons die sydigheid wat ontstaan as ons sommige van die simulasie aannames verander

of die data verander word om te verseker dat dit in ‘n vorm is wat bruikbaar is vir elk van die twee metodes. Ons vind dat die prestasie van die metodes afhang van die aannames wat die aan data-genereringsproses onderliggend is, die mate van ruimtelike samestelling, die gekose metode om die data te verander om dit in ‘n bruikbare vorm te kry vir die skattingsmetode en die parameter wat gepas word. Die S_0 beraam deur die verwydering metode op area vlak is byna onsydig wanneer die bevolking naïef is, maar sydig wanneer dit op die bevolkingsvlak geag word, of die bevolking aanvanklik naïef is of nie. Verder is die verwyderings metode $\mathcal{R}_0$ en $\mathcal{R}_{eff}$ ramings oor die algemeen sydig. Die TSIR-model, aan die ander kant, lyk beter om beide S_0 en $\mathcal{R}_0$ vir nie-naïewe bevolkings te beraam. Hierdie bevindings is nuttig omdat hulle ons ’n idee gee van die sydighede in die pas van hierdie metodes tot werklike data van dieselfde aard as die gesimuleerde epidemies. Vir die tweede vraag, beraam ons die impak van ‘n uitbraakrespons-inentingsveldtog wat georganiseer is in reaksie op ‘n maselsuitbraak in ‘n hoërskool in Stellenbosch, Suid-Afrika. Ons bereik dit deur ‘n diskrete stogastiese vatbaar-blootgestel-aansteeklik-herstel (SEIR) model met daaglikse tydstappe te formuleer, wat geboortes en sterftes ignoreer. Deur die model te gebruik, analiseer ons verskeie scenario’s wat ons toelaat om die aantal afgeweerde gevalle te skat, en om die oorblywende aantal gevalle te tot die epidemie geëindig het te skat en die tydsraamwerk waarbinne sulke gevalle sou plaasvind. Opsommend oor scenario’s, skat ons dat ‘n mediaan van 255 gevalle (omvang 60 - 493) afgeweer is. Daar is ook ’n mediaan van 15 oorblywende gevalle (omvang $1 - 33$) en ‘n mediaan van 4 oorblywende weke (omvang $1 - 16$) verwag totdat die epidemie geëindig het. Ons kom tot die gevolgtrekking dat die veldtog suksesvol was om baie potensiële gevalle te voorkom.

Acknowledgements

First, my utmost thanks goes to the God whom I serve. He has overwhelmed me with His grace, favour, and guidance, and has fulfilled His promises to me. Without Him, I would have caved in during my burnouts.

My supervisor, Prof Juliet Pulliam, is such a phenomenal scholar. Until I met her, I had not met a well-rounded scholar who has the eye for detail to such a nicety. Without her amazing time management skills, she would not have had enough time off her duties as Director of SACEMA to offer me every little explanation and guidance I needed to come up with this work. I hope to imbibe portions of her work ethics as she continues to supervise me in my future work.

Further thanks goes to Dr Gavin Hitchcock for his emotional support and constructive feedback on some aspects of my work. Apart from my academic work, he always made sure I was on top of my game in terms of maintaining a spiritual, physical and mental well-being. God bless him.

My late mom, Mrs Gladys Animai Azam, and my dad, Mr. Lawrence Akudugu Azam broke their last sweats to make sure I always got everything I needed to study without worry. I could not have gotten to this point if they were not there with me and guided me through my every decision. Mom was always worried to hear I was travelling far from home to study but I always assured her I would be fine. Today, I am fine, but she is not here to give me a hug of congratulations. May she rest in peace.

More thanks goes to SACEMA for the funding, and all my colleagues at SACEMA for the emotional support and line editing of my thesis. They were all amazing and contributed significantly to my success.

My special thanks goes to Zinhle Mthombothi, with whom I burnt the midnight oil.

The list is long, but I have limited space to express thank everyone. To any of my friends not specifically mentioned here, know that your thank you's are boldly written in a better place - my heart.

Dedications

I dedicate this work to my late mom, Mrs. Gladys Animai, who left us without warning, and my dad, who wore himself out (with mom, of course) to make sure I achieved my dreams. Ultimately, this goes to my family and all my friends. I did this to prove you can do it too!!!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Measles is a directly transmitted viral infection affecting only humans. Individuals who have never been infected, and have not been successfully vaccinated, are at risk of acquiring the disease. Of all human diseases, measles is among the most contagious, making it a public health issue because it is difficult to control. In fact, with as low as 5% of a population being susceptible, outbreaks are bound to occur (Orenstein & Gay, 2004).

Before the advent of a safe, affordable, and highly effective vaccine, measles caused over 2.6 million annual deaths among children across the world with most of the deaths emanating from developing countries. In 2008, measles was ranked as the tenth killer of children under 5 years by a systematic review of causes of under-five deaths (Black *et al.*, 2010). Even though the mortality rates have declined, measles continues to cause substantial deaths among children in developing countries, with almost 10- to 100-fold higher mortality rates compared with developed countries (Cairns *et al.*, 2011). These high rates of mortality led the World Health Assembly to set a mortality reduction goal in 2005 to reduce measles-related mortality by 90% by 2010, compared with the 2000 estimate of 544 200 (CI: 364 300 – 891 500) (Perry *et al.*, 2014). This goal was gazetted in the Global Immunization Vision and Strategy document (WHO, 2005).

In 2010, the goal was renewed to target a 95% reduction in measles-related mortality to be achieved by 2015. Two other goals accompanied the quest to reduce measles-related mortality: an increase in national and district-wide routine vaccination coverage of the first dose, reaching at least 90% of infants by their first birthday; and a reduction in annual global measles reported cases to below 5 cases per million population. These three goals were set to be achieved by 2015 but have not yet been reached.

Since the measles vaccine became readily available, recognizable strides have been made at reducing measles-related deaths. The vaccine is estimated to have prevented

over 20.3 million deaths. An annual decrease from 651 600 to 134 200 deaths between 2000 – 2015 has been estimated, representing a percentage decrease of approximately 79% (WHO, 2017). The progress made with the vaccine so far has partly been attributed to the Measles & Rubella Initiative (M&R Initiative), a global partnership comprising the American Red Cross, UNICEF, Centers for Disease Control and Prevention (CDC), WHO, and the United Nations Foundation. This partnership was established in 2001 with the mandate to eliminate measles and rubella in a minimum 5 out of 6 WHO regions by 2020. Reducing measles-related mortality is one of the M&R core goals.

The World Health Organization (WHO), in agreement with its six regions, has set clear measles objectives - elimination or mortality-reduction. The conviction that measles can be eliminated stems from the biological characteristics of the virus: humans are the sole reservoir, diagnostic tests are highly accurate at detecting the disease, and the vaccine is safe, effective and affordable. Even though the global goal is elimination by 2020, mortality-reduction is the immediate goal in the sub-Saharan Africa region, where incidence and mortality continue to peak. Interestingly, between 2006 – 2008, the Southern Africa region achieved elimination by reducing measles incidence to less than one case per million population but that was short-lived as large outbreaks occurred between 2009 – 2010 (Shibeshi *et al.*, 2014). This goes to prove that even though the objectives are clear, they are not easy to accomplish. As the year 2020 approaches, elimination success seems bleak in several of the WHO regions as immunity gaps continue to exist, resulting in continual outbreaks the world over. This is partly because, in populations with high vaccination uptake, the average age of infection shifts from under 5 years to the teen ages as a result of vaccine failure, or waning in some individuals. As the susceptible populations build up, they become hot spots for measles outbreaks. Immigrants and individuals who deliberately reject the vaccine are another source of outbreaks as well.

It has become increasingly evident that measles can only be eliminated through effective and efficient vaccination; vaccination helps to ensure high levels of population immunity, and breaks measles transmission (Moss & Griffin, 2012). It is therefore essential to ensure proper implementation of routine and second dose vaccination schedules targeting broader age ranges among children. This schedule should be supported with regular supplementary immunization activities for catching up those who missed the earlier campaigns. Reactive vaccination campaigns during outbreaks have been shown to be effective in curtailing measles epidemics and improving population level immunity (Cairns *et al.*, 2011) but they have to be timely (Grais *et al.*, 2007), and accompanied by: case management, to reduce mortality and measles-related complications (Klepac *et al.*, 2012); and improved surveillance, to detect new cases promptly (WHO, 2009). These

steps will ensure that immunity gaps are bridged in the quest to eliminate measles.

Mathematically, it is possible to infer from previously reported measles epidemics whether a country or region is close to measles elimination or not. This is achieved by making inferences about the transmission potential of measles using either of two quantities: on the one hand, we have the basic reproduction number $\mathcal{R}_0$, a quantity that measures the mean number of secondary cases a single case will produce in an entirely susceptible population (Anderson *et al.*, 1992; Orenstein & Gay, 2004). $\mathcal{R}_0$ is easy to interpret: if it is less than 1, the disease will go extinct; if greater than 1, the disease is likely to spread. $\mathcal{R}_0$ is ascribed to populations with no immunity against measles. On the other hand, we have the effective reproduction number $\mathcal{R}_{eff}$, a summary measure which describes the number of new cases likely to be produced by a single case in a population with some level of immunity. $\mathcal{R}_{eff}$ is used with populations where certain individuals are immune due to previous exposure (as is the case in most sub-Saharan African countries where the disease is either endemic or re-emerges irregularly) or where there are immunity gaps resulting from poorly implemented vaccination programmes. To eliminate measles, $\mathcal{R}_{eff}$ must be maintained below 1. In essence, $\mathcal{R}_0$ and $\mathcal{R}_{eff}$ have proven their utility in designing measles elimination strategies, and for assessing the impact of vaccination campaigns (Orenstein & Gay, 2004).

Being able to mathematically infer measles transmission potential from epidemic data using $\mathcal{R}_0$ and $\mathcal{R}_{eff}$ has been facilitated by the development of two models namely, the time series susceptible-infected-recovered (TSIR) model (Bjørnstad *et al.*, 2002) and removal model (Ferrari *et al.*, 2005). Whiles the TSIR model depends on having many years of epidemic data to infer the time-varying transmission potential $\mathcal{R}_0$ of a pathogen as well as the initial size of the susceptible population S_0, the removal model only works with short time series having a single peak, estimating a constant $\mathcal{R}_{eff}$ and the size of the susceptible population at the onset of the epidemic S_0. These models have been applied in a number measles outbreak settings and for studying vaccination impact (Ferrari *et al.*, 2014, 2008; Grais *et al.*, 2007, 2006*b*). In fact, an R (R Core Team, 2017) package has even been developed for implementing the TSIR model for fitting epidemic data (Becker & Grenfell, 2017), demonstrating the utility of the TSIR.

The TSIR and removal model have mostly been applied to well-curated datasets. In practice, however, some epidemic datasets are poorly collected and sometimes do not match the exact time scale of these two models nor all of the simplifying assumptions. A typical example of poorly curated data would be one containing unexplained missing data. It is simply not enough to remove the missing data points since any estimates obtained will be biased and may lead to an over- or underestimation of the actual trans-

mission potential of measles for that epidemic. In such situations, it is important to know what measures can be put in place to cut down on the biases while applying the models for fitting.

This thesis, therefore, focuses on the application of mathematical and statistical techniques to understanding measles model parametrization, and measles dynamics and control, in two African settings: First, we will assess the performance of the TSIR and removal method, using simulated data resembling poorly-curated datasets. We will interpret the models' performance on fitting measles data with particular types of errors that reflect those commonly found in countries with limited resources of data collection and curation. Hence, the aim here is to assess how the estimates compare with the actual value. Our findings will be important because they will give us an idea of the confidence to attach to the predictions resulting from the model fits of an actual poor quality data set in our possession. Second, we will assess the impact of a reactive vaccination campaign which was organised in response to a measles outbreak which occurred in a school in Stellenbosch, South Africa in February 2017. The outbreak, coupled with the response, provides an opportunity for us to study the progression of the epidemic following vaccination. The findings of Chapter 3 will help enrich the literature on the impact of outbreak response vaccination campaigns. The assessment will be done by determining how many cases were averted, the remaining time until elapse of the outbreak, and cases to expect.

In summary, this thesis will seek to address the following questions:

1. How well do the time series susceptible-infected-recovered (TSIR) model, and the removal method perform in predicting actual parameters from measles data when these parameters are known?

2. How many cases were averted following an outbreak response vaccination in a school in South Africa? Also,

    - how much time remained after the campaign till the epidemic elapsed?
    - how many more cases were expected until the epidemic died out?

To answer the questions previously listed, we will set out to:

1. fit simulated datasets with the TSIR model, and the removal method for comparison with the actual values

2. formulate a measles susceptible-exposed-infected-recovered (SEIR) model for a closed population

3. determine the number of cases averted, and cases and time remaining, following a vaccination campaign in the Stellenbosch school outbreak

This thesis is structured as follows: it begins with a general review of the related literature, followed by Chapter 2 where we assess the performance of two parameter estimation methods using simulated data. Chapter 3 investigates the impact of an outbreak response vaccination campaign in a school in South Africa. Chapter 4 draws the curtain on this thesis by briefly discussing the implications of our findings.

In the next section, we will be joining in the ongoing scholarly conversation about measles models and methods for their parametrization, and measles outbreak response vaccination. We will indicate how this thesis fits into this conversation.

1.1 Background to measles

Measles is a single-stranded RNA virus in the genus *Mobilivirus* and family *Paramyxoviridae*. Humans are the only known reservoir of the measles virus even though it is possible to infect some non-human primates - macaques specifically - in experimental settings (De Swart, 2009).

Measles is spread by direct contact with the aerosol droppings and phlegm of infected individuals (Griffin & Oldstone, 2008; Heymann, 2004; WHO, 2017), or with articles freshly contaminated by infected individuals (Heymann, 2004); the virus is known to remain active for up to 2 hours outside of the human body.

The natural history of measles has been well-described: The first stage is an asymptomatic incubation period of about 10 days, but varies among individuals. Generally, estimates of the incubation period lie between $7 - 23$ with an average of 10 days (Fine, 2003; Heymann, 2004; WHO, 2017). The incubation period is protracted when exposed individuals are administered with immune serum globulin, or the vaccine, as post-exposure prophylaxis. Following the incubation period, the individual exhibits a prominent rash for about $4 - 6$ days, accompanied by a fever, cough, sneezing, and reddening of the eyes. When the rash appears, it implies the individual has been infectious for an average of 4 days and will remain infectious for 4 more days (Griffin & Oldstone, 2008; Heymann, 2004; WHO, 2017). After the incubation and infectious stages, measles confers lifelong immunity on the individual (Blake & Trask Jr, 1921; Krugman *et al.*, 1965; Markowitz *et al.*, 1990).

When a person is suspected of having measles, it is advisable to use the World Health Organization (WHO) clinical case definition for diagnosis. The WHO case definition helps reduce misclassification because other diseases can be mistaken for measles, in-

cluding: rubella, parvovirus B19, human herpes virus 6 and 7, dengue virus, or scarlet fever. According to the definition, the initial symptoms are a fever accompanied with a prominent maculopapular rash. Individuals exhibiting such symptoms should be clinically tested to ascertain the measles infection with one of two test options: the immunosorbent assay (ELISA) test, which is used to probe for anti-measles virus immunoglobulin (IgM) antibodies; or the reverse transcriptase polymerase chain reaction (RT-PCR) which is used to test for measles RNA in throat swaps, urine, nasal mucus, or mouth fluids (WHO, 2017).

A number of factors put individuals at risk of acquiring measles. All children and adults without prior history of infection, and those who failed to develop an immune response from vaccination are susceptible. The measles vaccine reduces the risk of being infected by providing immunity but there is ongoing debate about whether this immunity is lifelong; some studies have suggested the possibility of waning immunity in vaccinees, hence there is the likelihood that vaccinated individuals become susceptible later in life (Glass & Grenfell, 2003; Heffernan & Keeling, 2008; Mossong & Muller, 2003; Mossong *et al.*, 1999). Newborns present an interesting case of susceptibility: if a mother has had measles before, she will transfer maternal antibodies to her newborn. Hence, for a period of about 9 months after birth, these babies are immune, but become susceptible when the measles antibodies wane. Additionally, mothers with vaccine-induced immunity transfer a lower concentration of antibodies to their babies, leading to quicker antibody waning. Consequently, these babies become susceptible earlier than the former group described. During the period of passive immunity, they are highly likely to fail vaccination (Krugman *et al.*, 1965; Moss & Griffin, 2012).

Measles causes a number of complications. The severity in measles symptoms depends on the strength of an individual's immune system and other factors including vitamin A deficiency, dehydration, undernourishment, and so on. In HIV-infected individuals, these complications are facilitated by the HIV virus which compromises the individual's immune system (Orenstein *et al.*, 2004; WHO, 2017). A review by Orenstein *et al.* (2004) indicates that measles affects a number of organs in the body, in the form of: otitis media, and pneumonia in the respiratory tract; encephalitis in the neurological system; diarrhoea in the gastrointestinal tract; blindness in the ophthalmic organs; complications in the cardiovascular system; rashes on the skin; and so forth.

Measles is incurable. Palliative care is provided to infected individuals to ease some of the initial symptoms like coughing, fevers, reddening of the eyes, et cetera. Additionally, measles cases are given oral rehydration to reduce dehydration, nutritional support is boosted, and medications are administered to reduce complications if any

arise. Vitamin A supplements are provided to boost the immune system. Antibiotics are not recommended. Palliative care is efficient in reducing measles mortality rates. Further, within 72 hours after being exposed to measles, the live-attenuated measles vaccine is sometimes employed as a post-exposure prophylaxis as it is known to ameliorate measles symptoms, causing a shorter duration of illness. However, since pregnant women, immuno-compromised individuals, and newborns less than 6 months are contra-indicated candidates for the measles vaccine, human immune globulin serves as a substitute to the vaccine within 6 days (Fulginiti, 1965; Heymann, 2004; Orenstein *et al.*, 2004; WHO, 2017).

Before a safe vaccine became available in the 1960's (Enders *et al.*, 1962), measles control practices included: quarantining infected individuals to decrease mixing with susceptible individuals; and administration of convalescent serum to highly compromised individuals to attenuate the symptoms, while allowing the disease to take its course. Additionally, bed-spacing was increased in sick bays in affected boarding schools to reduce spread (Hobson, 1938). Without a vaccine, almost all susceptible individuals acquired the infection.

Currently, the live-attenuated vaccine is safe, affordable and effective for controlling measles epidemics. The measles vaccine either comes combined with live-attenuated Mumps and Rubella vaccines (MMR), and sometimes, Varicella vaccines (MMRV), or as a monovalent measles-containing vaccine (MV). The most common option is the MMR. A single measles vaccine dose (MV1), generally administered by the first birthday or before 15 months, successfully immunizes susceptible infants 94% to 98% of the time. The first dose is given routinely in most countries. A second dose (MV2) is advised as a booster for those who have received the first, or as a catch-up for those who missed or failed the first. In low resource-settings especially in sub-Saharan Africa, it is difficult to implement the second round of doses. This second dose is, therefore, mostly provided through supplementary immunization activities (SIAs). In the developed countries like the United States and the United Kingdom, administration of the second dose is routinely implemented and targets individuals about to enter school ($4-6$ years) even though it can be received a few weeks after the first dose. SIAs in most developed countries are organised every $2-4$ years to serve as a second opportunity for those who missed the second dose. Also, in the event of measles outbreaks, reactive vaccination campaigns are organised targeting all individuals at risk (Enders *et al.*, 1962; Griffin & Pan, 2009; Heffernan & Keeling, 2008; WHO, 2017).

Coverage is the most important aspect of measles control when using the vaccine. In order for countries to achieve measles elimination, it is essential that vaccination cov-

erage levels are as high as possible, ideally above 95%. A high coverage ensures that unvaccinated individuals are protected. This is called herd immunity (Anderson *et al.*, 1992). Measles coverage is directly proportional to closeness to measles elimination. The closer the coverage is to 100%, the higher a country's chances are of eliminating the disease. Most immunization programs, therefore, have the objective of increasing coverage. Achieving high vaccination coverage is, however, hindered by certain brackets of the population who: deliberately reject the vaccine , are contraindicated for the vaccine, or are hard-to-reach.

Numerous measles models have centred on studying the impact of vaccination strategies on increasing coverage. These models are mostly developed on an underlying simple model. The next section will discuss a selection of models that have been used to study measles dynamics.

1.2 Models of measles dynamics

Much progress has been made in mathematically modelling the dynamics of measles transmission using various types of differential equations. Three of the most popular types of equations used include: ordinary and partial differential equations (ODE/PDE), for continuous time; and difference equations (DE), for discrete time. These models are generally population-based and assume that individuals can be compartmentalised according to disease state characteristics - susceptible, infected, or immune.

One model framework groups individuals who are at risk of acquiring the infection into the susceptible class (S), infectious individuals into the infected class (I), and individuals with immunity against measles into the removed class (R). This class of models are the susceptible-infected-recovered (SIR) models (Anderson *et al.*, 1992; Keeling & Rohani, 2008).

Another set of models modifies the (SIR) models by splitting the (I) class into two sub-classes: those who are infected but uninfectious - the exposed class (E); and those who are infectious - the infectious class (I). These class of models are the susceptible-exposed-infected-recovered $(SEIR)$ models.

The basic ordinary differential equation representation of the SEIR model is

$$\begin{aligned}\frac{dS}{dt} &= \nu - (\beta I + \mu)S \\ \frac{dE}{dt} &= \beta IS - (\sigma + \mu)E \\ \frac{dI}{dt} &= \sigma E - (\gamma + \mu)I \\ \frac{dR}{dt} &= \gamma I - \mu R\end{aligned} \tag{1.2.1}$$

where S, E, I, and R have earlier been described. The parameters ν and μ are the constant birth and death rate respectively, and β is the time-invariant transmission coefficient (Anderson *et al.*, 1992; Keeling & Rohani, 2008). In this model, the duration of latency and infectiousness are constant with a mean of $\frac{1}{\sigma}$ and $\frac{1}{\gamma}$ respectively.

Equations 1.2.1 have been the underpinning model for many discussions around measles dynamics. The SEIR model has been used to explore the dynamical behaviour of measles by investigating the influence of several factors such as vaccination uptake and the size of the population on the progression of the epidemic. In their study, Earn *et al.* (2000) argued that the $SEIR$ model, represented by Equation 1.2.1, is unable to capture complicated phenomena because it assumes a constant rate of susceptible recruitment (births). They drew insights from bifurcations of the mean transmission parameter by observing the interaction between birth and vaccination rate, and the mean transmission parameter to demonstrate that these dynamic transitions are explained by a seasonality in birth rate and vaccination uptake - dynamically, increased vaccination uptake translates as reduced susceptible recruitment, hence an interaction between the two parameters. Further, the birth and vaccination rate influence the rate of transmission so that the resultant is an effective quantity which drives measles spread.

Earn *et al.* (2000) also argued that the traditional method of introducing transmission seasonality with sinusoidal forcing is unrealistic since it is well known that measles transmission peaks during school terms and drops during off-school seasons (Bjørnstad *et al.*, 2002; Dorélien *et al.*, 2013; Fine & Clarkson, 1982). The study by Earn *et al.* (2000) shows that the defining element of $SEIR$ model inference is not in explicitly incorporating various levels of heterogeneities but rather, employing a realistic seasonal forcing function with a simplistic model. In contrast to this, an earlier study by Bolker & Grenfell (1993) had suggested that an SEIR model with age and spatial structure, and biological complexity would be the most basic formulation to describe non-linear measles dynamics.

A number of studies have employed modified SEIR-type models to infer the impact

of various measles outbreak control strategies. In this work, we only focus on vaccination.

In order to be useful, models should be appropriately parametrized. The SEIR model requires realistic estimates of: the transmission rate - and an idea of its seasonally-forced nature; the size of the population at risk - mainly because only the infections are reported; and the transmission potential of the measles pathogen within the population being studied. The next section examines the parameters relevant to this thesis' objectives and the techniques for estimating them.

1.3 Parametrizing measles SEIR models

In a completely susceptible population, the average number of secondary measles cases that a single case would produce is called the basic reproductive ratio, denoted as $\mathcal{R}_0$ (pronounced R nought) (Anderson *et al.*, 1992; Grais *et al.*, 2006*b*; Keeling & Rohani, 2008). $\mathcal{R}_0$ is central in epidemiology and serves as the threshold for determining: if a disease will spread $(\mathcal{R}_0 > 1)$ or die out $(\mathcal{R}_0 < 1)$; the herd immunity arising from vaccination; the vaccination coverage required to achieve herd immunity (Guerra *et al.*, 2017); the final epidemic size - number of people infected at the end of the epidemic; and the mean age of infection (Anderson *et al.*, 1992; Grais *et al.*, 2006*b*).

The effective reproduction number $\mathcal{R}_{eff}$, defined as the number of secondary measles cases produced by a single infected case in a partially susceptible population (Grais *et al.*, 2006*b*), is used for inferring measles transmission potential from populations with a previous record of outbreaks or vaccination. $\mathcal{R}_{\text{eff}}$ is useful for determining: the epidemic duration, the timing of epidemic peaks, the least number of susceptible individuals to be removed from the population to ensure herd immunity is attained, and the impact of an immunization campaign (Grais *et al.*, 2006*b*).

A direct relationship exists between the basic reproductive ratio and the effective reproduction number, that is, $\mathcal{R}_{\text{eff}} = s\mathcal{R}_0$, where s is the proportion of the population that is susceptible (Anderson *et al.*, 1992).

In recent literature, two methods for estimating $\mathcal{R}_0$ and $\mathcal{R}_{\text{eff}}$ have been presented: the TSIR model by Bjørnstad *et al.* (2002), and the removal model by Ferrari *et al.* (2005). Both methods are essentially chain binomial models, and they are described in detail in Chapter 2. Basically, the removal model by Ferrari *et al.* (2005) draws insights from the removal method in ecological modelling, making simplistic assumptions - homogeneous mixing and no susceptible recruitment - to estimate $\mathcal{R}_{eff}$, and S_0 from epidemic data, having a single peak. The time series susceptible-infected-recovered (TSIR) model,

developed by Bjørnstad *et al.* (2002), is based on the assumption that mixing is heterogeneous and that susceptible individuals are recruited through birth. This model also requires a long time series of measles case counts spanning multiple years. The added advantage of the TSIR model is that it can estimate a seasonally changing transmission rate which is useful for reproducing complicated measles dynamics.

Measles epidemics are often well-described, and documented, and have been a rich source of information for advancing the theory of dynamic modelling of infectious diseases. Measles data have, in particular, been employed in various studies for validating parameter estimation techniques (and models) like the TSIR and removal model. Detailed records exist for: epidemics in large cities in England, Wales, United States and Denmark from the pre-vaccination era, described in Bjørnstad *et al.* (2002); Bobashev *et al.* (2000); Finkenstädt & Grenfell (2000); Metcalf *et al.* (2009); a combination of pre- and post-vaccination era epidemics, for example, England and wales described in Fine & Clarkson (1982); Pre-vaccination era epidemics in schools, such as those described in Hobson (1934, 1938); epidemics in highly-vaccinated schools, one of which is described in Yeung *et al.* (2005); and epidemics in cities with uneven vaccination coverage in developing countries, two of which are described in Grais *et al.* (2006*a*,*b*).

In this Chapter 2, we assess the removal method and the TSIR model using simulated data resembling that collected in low-income countries where data usually contains high levels of reporting bias and data collection errors .

1.4 Measles outbreak response vaccination in low-resource settings

Measles outbreak response vaccination/immunization (ORV/ORI) is broadly defined as any vaccination campaign, other than routine campaigns, organized in reaction to an outbreak of measles in a country (Cairns *et al.*, 2011). The definition of an outbreak is context-specific and depends on the country's measles control phase, that is, their closeness to elimination, and whether the country has previously undertaken a supplementary immunization activity (SIA) or not (WHO, 2009). In general, an outbreak is declared when cases emerge in a population in a manner meeting certain criteria such as how many cases have been laboratory-confirmed and how many weeks have seen successive measles cases.

In the past, the World Health Organization (WHO) discouraged ORVs as a first line of defence during outbreaks on the basis that these campaigns were generally delayed, yielding insignificant impact on preventing cases, and also led to unprofitable expendi-

ture (WHO, 1999). The guidelines, however, advised that in situations where campaigns were being considered, unaffected and high risk areas were to be prioritized, especially if there were high numbers of susceptible individuals that could serve as hotspots for rapid spread. This view was underpinned by a number of studies, one of which was undertaken by Aylward *et al.* (1997) who, in their review on papers published on the impact of outbreak response campaigns in middle and low income countries, concluded that ORVs should be discouraged since there was insufficient data to support their success and that the effort should rather be directed at outbreak prediction and prevention through routine and supplementary vaccination strategies and better surveillance. But this WHO stance has since been revised due to increasing evidence of the success of ORVs (Grais *et al.*, 2006*a*, 2007, 2006*b*; Lessler *et al.*, 2016). Recent WHO guidelines on outbreak response in low-resource settings emphasize reducing morbidity and mortality by managing cases and vaccinating children. The secondary goals of ORV in mortality-reduction settings include minimizing outbreak duration and cases; improving surveillance and population-level immunity; and raising awareness of the disease and preventive measures (WHO, 2009).

1.5 Models and methods for assessing outbreak response vaccination impact

A number of studies have defined and used various success metrics to assess the impact of outbreak response vaccination, including, following ORV: duration of outbreak, local and national persistence, distribution of the cases over time, cases averted, time spent, and percentage of actual susceptible individuals vaccinated (Cairns *et al.*, 2011; Minetti *et al.*, 2013).

In one such study, Grais *et al.* (2006*a*) analysed measles epidemic data from two African cities namely, Kinshasa (2002-2003) and Niamey (2003-2004) and found that the epidemic spread was slow in both cities and that a rapid immunization campaign would have prevented a considerable number of cases. They employed statistical analyses of attack rates, as prescribed by WHO (WHO, 2009), to describe the severity of the epidemic, and determined spatial spread to a new district based on the condition that cases were reported in that district within two consecutive weeks. Their study raised the question of whether the campaign would have been more effective if it had been conducted earlier and targeted a wider age range.

To answer the question raised, they followed up with a modelling study to estimate the impact of campaigns based on speed of response and wider age targets. Here, Grais

et al. (2007) formulated a stochastic $SEIR$ model with daily time steps, allowing for the tracking of individuals, to study the potential impact, in terms of cases averted, of the campaign organised in Niamey between $2003 - 2004$.

In their model, the process of infection was modelled assuming a binomial distribution with the number of susceptible individuals at any time S_t, as the denominator, and the associated probability of new infections as $1 - exp(\frac{-\beta I_t}{N})$, where β is the transmission coefficient, I_t is the number of cases at a particular time, and N is the total population. The transmission parameter was weighted so that those who lived closer to the infected individuals possessed a higher chance of getting infected than those who lived farther away. Also, in the model, the process leading to recovery, after being infected, was treated as deterministic with an incubation and infectious period of 10 and 6 days respectively. The model had a spatial component, allowing for children to be categorized into spatial groups comprising, in increasing area: quartiers (neighborhoods), communes (districts), Centre de Santés Intégré (health centers), and cities.

Using this model, they found that even though the outbreak response campaign was successful in preventing numerous potential cases, it would have been more efficient if the response was faster and targeted a wider age range. This result therefore agreed with the prediction of rapid response as an efficient outbreak response vaccination strategy in Grais *et al.* (2006*a*).

When modelling to assess the impact of outbreak response vaccination, it is important to incorporate realistic assumptions in the model in order not to over- or underestimate the impact. Spatial and demographic heterogeneities are two of such assumptions which should not be overlooked. The limitation, however, lies in the data available. It is only possible to incorporate such heterogeneities if the data allows for that.

In a more recent study, Lessler *et al.* (2016) adopted a time series susceptible-infected-recovered (TSIR) model (with age structure and seasonal forcing) from Metcalf *et al.* (2012) to study the impact of triggered campaign (TC) strategies on simulated measles outbreaks in four synthetic populations closely resembling countries making strides at achieving measles elimination. A triggered campaign here refers to any campaign, other than an assumed baseline routine campaign and 5-yearly supplementary immunization activity (SIA), initiated either following the detection of a certain level of susceptibility in a target age group through a sero-survey, or an "unusual" number of cases identified in the population.

The "countries" were chosen according to the nature of their birth rates and vaccination coverage: moderate birth rate and routine coverage; high birth rate, moderate coverage; moderate birth rate, high coverage; high birth rate and coverage.

In their simulation, they introduced a single case of measles into the population and tracked the epidemic for forty years taking note of how many: triggered campaigns occurred, cases were prevented through the campaigns, and total cases were produced by the single case.

They found that case-triggered campaigns averted more cases than the baseline and more markedly in populations where the baseline achieved low coverage, but serologically triggered campaigns were more effective in reducing the size of the epidemics compared to case-triggered campaigns or the baseline. The lesson to be learnt from this work is outbreak prevention through serological surveys may be more effective in reducing the disease burden than response campaigns.

A stronger argument could have been made by Lessler *et al.* (2016) if a simple cost analysis was performed to compare serological campaigns to ORVs. In low-resource settings the cost involved in conducting such surveys might discourage the authorities and cause them to lean more towards ORVs since, at face value, ORVs seem cheaper. The work by Lessler *et al.* (2016), even though somewhat simplistic and neglected a number of heterogeneities, paves the way for more comprehensive simulation studies to be done to influence local policies on ORV on a case-to-case basis as has been advised by WHO (WHO, 2009) and (Cairns *et al.*, 2011).

Generally, ORVs are non-selective with regard to the immune status of vaccinees, with a few exceptions (Cairns *et al.*, 2011). Non-selective campaigns usually target all children within the 9 months−15 years age range. However, Minetti *et al.* (2013) argue that this status quo risks the chance of re-vaccinating already immunized individuals, leading to a lower effective impact (or coverage) - how many susceptible individuals are vaccinated. Non-immunized individuals from previous campaigns are probably never reached due to limited access to vaccination services due to inaccessible road networks, distance from the vaccination sites, civil unrests, and so on. Yet, reported vaccination coverages do not seem to highlight the effective coverage achieved.

Therefore, Minetti *et al.* (2013) proposed effective impact as a better success metric than vaccination coverage, especially in countries with high population immunity. To substantiate their argument, they developed a stochastic SEIR model (with seasonally forcing), assuming 6 days of exposure and 8 days of infectiousness, and a non-naive population split into high- and low-access categories to estimate the effective impact, by way of coverage, of an ORV campaign organised in Malawi (2010). They found that the effective impact of the Malawi ORV was lower than expected even though the campaign achieved a high overall coverage of 95%. They were able to show that in high coverage populations, selective campaigns prioritizing hard-to-reach individuals would be more

beneficial in the journey towards elimination.

As Lessler *et al.* (2016) noted, not much is known concerning the impact of ORVs and more work needs to be done. Even though South Africa has made strides in reducing the number of outbreaks and cases as well as eliminating measles mortality and morbidity (Biellik *et al.*, 2002; Shibeshi *et al.*, 2014), it has not succeeded in completely eliminating the disease in fulfilment of the 2020 goal (Shibeshi *et al.*, 2014) which is fast approaching. Closed populations like schools provide a unique opportunity to study the impact of interventions as a contribution to the growing literature on model-based impact assessment of outbreak response vaccination. The outbreak in a highly vaccinated school in Stellenbosch-South Africa incited the need to draw some lessons towards measles elimination in South Africa and Southern Africa as a whole. The work done in Chapter 3 is the only one of its kind done for the region and opens the doors for more enquiry and refinement.

Chapter 2

Evaluating two methods for estimating measles parameters

2.1 Introduction

An initial step towards reproducing observed measles dynamics with a model is to obtain, from available data, estimates of important underlying biological processes like the rates of: transmission, birth, death due to disease, and recovery. This process, known as data fitting or model parametrization, is the contextualizing of models to a specific pathogen (and population) (see Keeling & Rohani, 2008: pg. 48).

In order to predict the behaviour of a measles epidemic following an intervention - like vaccination - it is important, first of all, to estimate the susceptible population size, that is, the population at risk of acquiring the infection. Knowing the size of the susceptible population is one step towards designing effective campaigns to efficiently target these groups. Besides, the susceptible population is generally unobserved during epidemics, hence knowledge of its distribution during an epidemic is useful in investigating the routes and dynamics of measles spread. This knowledge has direct implications for implementing measles control strategies, for instance, what proportion of the susceptible population to remove through vaccination to provide an overall protective effect - herd immunity (Bobashev *et al.*, 2000).

Generally, inferences on measles contagion are performed with either of two parallel quantities - $\mathcal{R}_0$ or $\mathcal{R}_{eff}$ - depending on whether the population possesses pre-existing immunity or not, and whether the immunity can be properly accounted for. In Chapter 1, we mentioned that estimating $\mathcal{R}_0$ is simple and straightforward for measles-naive populations, hence different methods of estimation have been developed in the litera-

ture. Studies like Cori *et al.* (2013); Metcalf *et al.* (2009); Wallinga *et al.* (2001), employed a variety of methods to estimate $\mathcal{R}_0$ from available data. It, however, becomes incrementally difficult to estimate measles reproductive potential from populations with pre-existing immunity, acquired actively or passively, because it is almost impossible to accurately gauge the exact immune proportion of the population without conducting a sero-survey. Notwithstanding, in partially immune populations, the effective reproductive number $\mathcal{R}_{\text{eff}}$, is used to make inferences pertaining to measles transmission potential. Many studies, including Chiew *et al.* (2014); Grais *et al.* (2006*b*), have estimated $\mathcal{R}_{eff}$ from measles case reports, using techniques they developed for the purpose.

Models for measles data fitting can broadly be categorized into discrete-time or continuous-time, depending on whether time is treated in the model as a continuous or discrete quantity. Both types of models have been employed successfully in modelling measles dynamics. The appealing nature of discrete-time models is that their time-steps can be made to match that of the reported data, allowing fairly easy parametrization, and tracking of individuals in various disease states. Two such discrete-time models are the time series susceptible-infected-recovered (TSIR) (Bjørnstad *et al.*, 2002) and removal estimator model (Ferrari *et al.*, 2005), which are basically chain binomial models.

The purpose of this chapter is to study the performance of each of these two estimation methods by carefully investigating how a mismatch between the underlying assumptions concerning the data - and the model assumptions - biases their estimates. By "performance", we mean we will be comparing the actual parameter values used to generate the stochastic epidemics, with the distribution of the fitted estimates to give us an idea of how well the methods perform when applied to actual data. Ferrari *et al.* (2005) evaluated the removal method by simulating data with the same model and estimating the parameters at each time step, which they compared with the actual parameter values. The model estimates matched the actual because the same model was used for both the simulation and fitting. The TSIR model has not been evaluated using our approach before, hence our findings will help shed more light on the confidence to associate with the TSIR estimates. Essentially, we want to know the biases that result when the data generation mechanism is different from that of the estimating model.

The rationale behind this chapter is simple. A number of assumptions underlie the derivation of the two parameter estimation methods; as a result, biased estimates are deemed to result when these assumptions are violated. As an example, the removal estimation method assumes that cases are perfectly reported (Ferrari *et al.*, 2005). It is known that measles cases are generally under-reported (Bobashev *et al.*, 2000; Fine & Clarkson, 1982). The authors acknowledged this as a limitation to interpreting the re-

sults of their model (see the discussion of Ferrari *et al.* (2005) for details). Similarly, the underlying chain binomial likelihood function for the removal method relies on previous infections to generate new ones and so, it is formulated to depend on a time series of non-zero cases. As a consequence, using data containing spontaneous generations of zeros, for estimation with this method, is problematic. Moreover, reported cases at the large spatial-temporal scales are usually greater than zero, but when the data is provided at a finer spatial resolution, some spatial components reveal spontaneous zero cases. As mentioned earlier, the removal method is based on the assumption that the time series cannot have zero cases. The TSIR model, on the other hand, assumes the whole population eventually gets infected in their lifetime, implying there is no prior immunity (Bjørnstad *et al.*, 2002). If that is not the case with the data being fitted, one common fix is to truncate the observations following vaccination, if any. This action is undesirable since the TSIR method requires long time series and so the remaining data might not be enough to produce desirable estimates. However, if the vaccination patterns are known, the model can be modified to incorporate them to avoid truncating the data.

The idea behind this chapter arose when we made initial attempts at estimating parameters from the data described in Section 2.2.1, with the two models. The data set breaks many of the TSIR and removal method assumptions. It became clear that both the TSIR and removal models fail when data, rife with missing observations and spontaneous zero cases, is fed to the models. Missing observations cannot be replaced or removed without a proper understanding of the reason for missingness. This warranted the use of a simulated data set, similar to our data set but having the advantage that we know the parameters used to generate them, to test the performance of the two models. Model performance can appropriately be measured if the actual parameter values being estimated are known, hence the use of simulated data in this case. We decided to assess the performance of the two models, after enacting some changes to the simulated data, mimicking what we intended to do with the actual data. The aim of this study was, therefore, to provide an idea of the influence of the biases resulting from artificial modifications to the data set that would allow the data to be usable by the two models.

This chapter is organised into four sections. The first section is an introduction, followed by the methods section in which we describe the data and mathematical derivations of the parameter estimation methods and the procedure for assessing the models. The third section is dedicated to reporting the results, after which we present our discussion of the results in the final section.

2.2 Materials and methods

2.2.1 The data

The Democratic Republic of Congo (DRC) is measles-endemic, so that epidemics occur regularly. During these epidemics, data are curated by the Ministry of Health in collaboration with Medicines Sans Frontieres (MSF) during outbreak response vaccination campaigns. For this study, the data available comprises a time series of weekly case counts spanning the period $2006-2011$ and 2015, for 438 health areas in the DRC - health areas are centres from which outbreak response vaccination campaigns are carried out in the DRC; several health areas make up a health zone. Currently, the data cannot be shown because of some restrictions by the provider.

Issues arising from our preliminary analyses of subsets of this data set inspired our primary question. We initially selected four focal health zones in the Katanga province, namely Malemba Nkulu, Lwamba, Mulongo, and Mukanga, with the aim of parametrizing a model we are developing for the Katanga context. It was revealed in the data cleaning process that observations for the period $2012-2014$ are completely missing with no documented reasons for the lack of information. Reported cases for the 2015 period, however, are complete with no missing observations. The $2006-2011$ time series is riddled with spontaneous missing data points especially at the health area level.

Additionally, at the health zone level, the data contains spontaneous zero cases which do not permit use with the TSIR or removal method. In order for us to carry out further analyses with the data, we needed to take important decisions about how to deal with the zero cases in the time series. We left the issue of dealing with the missingness in the data to future studies, and decided to use a simulation study to assess the performance of the parameter estimation methods when the zero cases had been replaced and assess the extent to which this will bias the estimates.

Before delving into how the model assessment was performed, we will first describe the TSIR and removal method in detail, and also provide the basis for the simulations used to generate the simulated data sets. We begin by describing the chain binomial model, which is the underpinning model for formulating the removal and TSIR models.

2.2.2 Chain binomial models

The chain binomial is a discrete time stochastic model. It derives its name from the idea that time series of infections at each time step of an epidemic can be generated through chains of binomially distributed processes, with the number of susceptible and infected

individuals in the population from the previous time step, S_t and I_t respectively, as input parameters for the binomial distribution (Bailey, 1975; Ferrari *et al.*, 2005).

Mathematically, the model can be explained as follows: Let Δt be a specified time step, such as the mean infectious period (Ferrari *et al.*, 2005). We denote, at time t, the number of susceptible individuals by S_t, and infectious individuals by I_t. Also, let the transmission parameter be β.

The parameter β is inherently the product of two underlying quantities: the rate of contacts a susceptible individual makes with infected individuals, and the probability that those contacts lead to a successful transfer of infection in the time interval Δt. Therefore, the transmission parameter represents the rate of contacts which lead to a successful transfer of the pathogen between two individuals within the given time step.

The probability of an individual escaping an infectious contact during the time period Δt is $e^{-\beta I_t \Delta t}$. Hence, new infected individuals are generated from existing susceptible individuals according to

$$I_{t+\Delta t} \sim \text{binomial}(S_t, 1 - e^{-\beta I_t \Delta t}). \tag{2.2.1}$$

The probability of producing κ infectious individuals at time t is therefore modeled as

$$P(I_{t+\Delta t} = \kappa) = \binom{S_t}{\kappa} (1 - e^{-\beta I_t \Delta t})^{\kappa} (e^{-\beta I_t \Delta t})^{S_t - \kappa}, \tag{2.2.2}$$

and the remaining susceptible individuals are generated from an iterative equation given by

$$S_{t+\Delta t} = S_t - I_{t+\Delta t}. \tag{2.2.3}$$

The chain binomial model is Equation 2.2.1 and 2.2.3. The model has several underlying assumptions: the pathogen being modelled has a constant latent period and a short infectious period, and an infected individual becomes immune after the course of infection. The model's stopping criterion is satisfied when the population runs out of susceptible individuals, or when there are not enough infectious individuals to protract the epidemic (Bailey, 1975).

Next, we will discuss the removal estimator model in detail.

2.2.3 Removal estimation method

The removal estimation method estimates $\mathcal{R}_0$, and the susceptible population at the onset of the epidemic S_0, from a time series of case counts. It is built on the chain binomial

model with the maximum likelihood framework as the underlying principle for parameter estimation.

The method has several advantages over a number of the other $\mathcal{R}_0$ estimation approaches. First, the model's $\mathcal{R}_0$ estimate is bias-corrected to account for any shortfalls resulting from approximating a continuous time birth-death process with a discrete time model (chain binomial), especially - this correction is only valid when we assume an exponentially distributed infectious period (Ferrari *et al.*, 2005). Epidemic data usually lacks information on the actual susceptible population. The removal method, nevertheless, provides an estimate of the initial susceptible population S_0, through a method of susceptible reconstruction proposed by Bobashev *et al.* (2000), and Finkenstädt & Grenfell (2000). The estimate of initial susceptible population is useful for predicting the susceptible dynamics as well as calculating the initial immune proportion in the population. Furthermore, the method is deemed to produce robust estimates, even though it assumes that individuals have equal contact probabilities, making random contacts irrespective of age, location, or behaviour. Additionally, the method requires little data (that is, an epidemic with a single peak only) to produce estimates and can be used in real-time while an epidemic progresses (Ferrari *et al.*, 2005). The usefulness of R_0 in epidemiology is undebatable. Hence, the ability of the removal process to estimate this quantity in real time is extremely useful especially for an emerging pathogen. Other methods of parameter estimation for instance, the time series susceptible-infected-recovered (TSIR), which we will discuss in the next section, require multiple years of epidemic data which is often either unavailable, or obtainable but with issues of under-reporting and poor quality.

We will now proceed to discuss how the removal method is mathematically derived.

2.2.3.1 Formulating the removal method

The mathematical formulations presented here follow that in the original paper by Ferrari *et al.* (2005).

We begin by recalling, from Section 2.2.2, the probability of producing κ infectious individuals at any given time t, written as

$$P(I_{t+\Delta t} = \kappa) = \binom{S_t}{\kappa} (1 - e^{-\beta I_t \Delta t})^{\kappa} (e^{-\beta I_t \Delta t})^{S_t - \kappa}.$$

In addition to that we recall Equation 2.2.3, which we rewrite as

$$S_t = S_0 - \sum_{i=1}^{t} I_i, \tag{2.2.4}$$

after introducing summations on both sides and shifting the indices.

With the equations above, we can proceed to set up the conditional probability for generating the time series of infections, which depends on the initial susceptible population S_0, and the transmission parameter β, as

$$P(I_{t+\Delta t}|I_t, I_{t-\Delta t}, \ldots, I_0, \beta, S_0) = \binom{S_t}{I_{t+\Delta t}} (1 - e^{-\beta I_t \Delta t})^{I_{t+\Delta t}} (e^{-\beta I_t \Delta t})^{S_t - I_{t+\Delta t}}, \tag{2.2.5}$$

where $t = 0, 1, 2, \ldots, T$.

The time step Δt in Equation 2.2.5 is the mean infectious period $\frac{1}{\gamma}$, being the inverse of the recovery rate γ. What this means is that if an individual becomes infectious at time t, then Δt is the time interval $(t, t + 1/\gamma)$, over which the individual remains infectious.

Here, we will substitute equation 2.2.4 into equation 2.2.5, while rescaling time to be in steps of the mean infectious period, so that we have

$$P(I_{t+1}|I_t, I_{t-1}, \ldots, I_0, \beta, S_0) = \binom{S_0 - \sum_{k=1}^{t} I_k}{I_{t+1}} (1 - e^{-\beta I_t})^{I_{t+1}} (e^{-\beta I_t})^{S_0 - I_{t+1} - \sum_{k=1}^{t} I_k}. \tag{2.2.6}$$

So far, we have set up the conditional probability expressing the new generation of cases in terms of the time series of previous cases $I_0, \ldots, I_t$, the initial susceptible population size S_0, and the transmission parameter β.

To estimate β and S_0, we use the method of maximum likelihood (MLE). Therefore, we need to write out the explicit form of the likelihood function. By employing Equation 2.2.6, the function is

$$\mathbb{L}(\beta, S_0|I_t, I_{t-1}, \ldots, I_0) = \prod_{t=1}^{T} P(I_t|I_{t-1}, \ldots, I_0, \beta, S_0). \tag{2.2.7}$$

The maximum likelihood estimates of S_0 and β, denoted as $\hat{S_0}$ and $\hat{\beta}$, are found by maximizing Equation 2.2.7, with the intrinsic assumption that the initial number of cases I_0, is 1, and that each new generation of infection is only dependent on the infectious individuals from the previous time step. The MLE of $\mathcal{R}_0$, $\hat{\mathcal{R}}_0$, is the product $\hat{S_0}\hat{\beta}$. This estimate, $\hat{\mathcal{R}}_0$, is calculated by following the parallel argument for computing $\mathcal{R}_0$ in the model as follows: the probability of acquiring infection at time $t+1$ is given by $1 - e^{-\beta I_t}$. Therefore, given S_t individuals mixing randomly with infected individuals at time t, we expect, according to the binomial distribution, $S_t \left(1 - e^{-\beta I_t}\right)$ new infections at time $t+1$. We denote the expected new infections by $\mathcal{R}_t$. Specifically, at $t = 0$, the number of new infections, $\mathcal{R}_0 = S_0(1 - e^{-\beta I_0})$. By assuming that the rate of transmission is small at the onset of the epidemic, the component $1 - e^{-\beta I_0}$ approximates β for $I_0 = 1$. This is true

because, generally, for small x, $1 - e^{-x} \approx x$. Therefore, $\mathcal{R}_0 \approx \beta S_0$, and $\hat{\mathcal{R}}_0 \approx \hat{S}_0\hat{\beta}$. In all of this, we have assumed that the population is constant, since infection and recovery are the only processes occurring in the model.

Finally, according to the removal estimation framework, $\hat{\mathcal{R}}_0$ is bias-corrected with a multiplicative factor of $2(1 - e^{-1}) - e^{-1}$. Details on how the discretization bias correction factor was derived, can be obtained from Section 2.1 of Ferrari *et al.* (2005). The resulting estimate, therefore, is

$$\hat{\mathcal{R}}_0 = \hat{S}_0\hat{\beta} \times 2(1 - e^{-1}) - e^{-1}$$

and after re-arranging,

$$\hat{\mathcal{R}}_0 = 2\hat{S}_0\hat{\beta}(1 - e^{-1}) - e^{-1}. \tag{2.2.8}$$

In the next section, we will discuss the TSIR model.

2.2.4 Time series susceptible-infected-recovered (TSIR) model

It has been shown, using measles data from England and Wales, collected during the periods before the introduction of measles vaccines that epidemics were either episodic or regular, depending on the size of the population, relative to the critical community size (CCS) - this is a measure of the population size required for an infection to persist without any importations (Ferrari *et al.*, 2008; Keeling & Grenfell, 1997). Large populations, that is, those above the CCS, exhibit regular epidemic cycles whiles small populations display irregular cycles interspersed with periods of local extinction.

The time-series susceptible-infected-recovered (TSIR) is a statistical model developed by Bjørnstad *et al.* (2002) to capture both episodic and regular epidemic dynamics. The TSIR model is an extension of an earlier model by Finkenstädt & Grenfell (2000), developed using a time series formulation. The Finkenstädt & Grenfell (2000) model is a mechanistic discrete time stochastic variant of the basic SEIR model developed to capture only the regular measles epidemic cycles exhibited by large communities but is not sensitive to small population sizes. In comparison, the TSIR model scales smoothly with population size. In particular, for small populations (below the critical community size), it is able to capture the resurgence of a disease after it has gone into local extinction, by accounting for repeated reintroduction of the pathogen (Bjørnstad *et al.*, 2002). The TSIR model functions by combining: (1) reconstructed susceptible dynamics, using the method by Bobashev *et al.* (2000); and (2) measles incidence data.

We shall proceed to discuss how the model is formulated.

2.2.4.1 Formulating the TSIR model

One underlying assumption of the TSIR model is that the time step is equivalent to the generation time - the sum of the incubation and infectious period - of the pathogen being studied. For measles, this time-step is roughly 2 weeks (a bi-week) (Anderson *et al.*, 1992). Therefore, in applying the model, data on reported measles cases are binned into bi-weekly intervals.

Another common assumption is that the transmission parameter β, varies bi-weekly, with a 1-year period of oscillation, denoted by β_s, where the subscript s represents the 26 bi-weeks in a year. Furthermore, the model assumes two sources of infection at any time interval t: infections due to importation, denoted as θ_t; and local infections, denoted as I_t, so that at every time-step there are $\theta_t + I_t$ infections, and hence, the force of infection τ, is given by $\beta_s(\theta_t + I_t)$.

Heterogeneities - example, in mixing - are captured by a parameter denoted α_1. A susceptible individual therefore experiences a force of infection at time interval t and bi-week s, given by

$$\tau = \beta_s(I_t + \theta_t)^{\alpha_1}, \qquad t = 1, 2, \ldots, T.$$

By assuming that the susceptible population at any time S_t, has a probability distribution such that its expected value is given by $E[S_t] = \bar{S}$, we can write $S_t = \bar{S} + D_t$, where D_t is the deviation around the mean susceptible population $\bar{S}$, where $E[D_t] = 0$ (Finkenstädt & Grenfell, 2000). Furthermore, the epidemic intensity that is, the expected number of new cases occurring one time step ahead λ_{t+1}, can be expressed as

$$E[I_{t+1}] = \lambda_{t+1} = \beta_s(I_t + \theta_t)^{\alpha_1}(\bar{S} + D_t).$$

Additionally, if we assume that susceptible individuals mix in a heterogeneous fashion, a second parameter α_2, is introduced into the previous expression for S_t, to obtain $S_t^{\alpha_2} = (\bar{S} + D_t)^{\alpha_2}$. Hence, the expected number of new cases is

$$E[I_{t+1}] = \lambda_{t+1} = \beta_s(I_t + \theta_t)^{\alpha_1}(\bar{S} + D_t)^{\alpha_2}. \tag{2.2.9}$$

Finally, the recursive equation for reconstructing the susceptible class, which is replenished by births at time t B_t, and expended by new infections I_{t+1}, is

$$S_{t+1} = B_t + S_t - I_{t+1}. \tag{2.2.10}$$

The TSIR model is represented by Equations 2.2.9, 2.2.10, and two other components depending on the assumptions made concerning the probability distribution of

local infections I_t, and the imported cases θ_t. Concerning the later two components, in the original TSIR formulation by Bjørnstad *et al.* (2002), and its sequel by Grenfell *et al.* (2002), demographic stochasticity in the epidemic birth-death process is assumed. Further, newly infected individuals are assumed to be produced in an independent process so that each distinct set of secondary cases is independently geometrically distributed. As a result, the dynamics of the infectious time series, being a series of independent geometrically distributed random variables, is approximated with a negative binomial distribution according to

$$I_{t+1} \sim NB(\lambda_{t+1}, I_t). \tag{2.2.11}$$

From Equation 2.2.11, I_t is the clumping parameter of the negative binomial distribution and λ_{t+1} is the expected number of new infections. Lastly, they assume that infections are imported at random (introducing an additional level of stochasticity into the model) so that the variable representing the count of immigrant cases follows a Poisson distribution, with a constant mean γ, according to

$$\theta_t \sim Poisson(\gamma). \tag{2.2.12}$$

They refer to equations 2.2.9, 2.2.10, 2.2.11, and 2.2.12 as a doubly stochastic formulation of the TSIR model (Bjørnstad *et al.*, 2002).

In the next section, we will describe one way of estimating the parameters.

2.2.4.2 Estimating the TSIR model parameters

A few methods have been employed to estimate the TSIR model parameters. For example, Bjørnstad *et al.* (2002), (Dalziel *et al.*, 2016), and Metcalf *et al.* (2009) used a regression technique after log-transforming Equation 2.2.9. In another study by Ferrari *et al.* (2008), they employed a Bayesian state-space approach to estimate the rate of under-reporting by assigning prior probability distributions to each parameter. We have chosen to use the regression technique because it is simple. It is described below, following the notation of Metcalf *et al.* (2009) and (Dalziel *et al.*, 2016).

Before we discuss the regression estimation framework, we need to reveal further assumptions which facilitate the estimation process. In that vein, the model assumes that the true cases are under-reported and moreover, with a constant reporting rate denoted by ρ. The true cases I_t, are therefore related to the reported cases $I_t^{(r)}$, by the equation

$$I_t = \frac{I_t^{(r)}}{\rho}$$

where $0 < \rho \leq 1$, and $\rho = 1$ implies the cases are perfectly reported.

Hence, when we make the above substitution into equation 2.2.10, we obtain

$$S_{t+1} = S_t + B_t - \frac{I_{t+1}^{(r)}}{\rho}. \tag{2.2.13}$$

Introducing sums on both sides of the equation,

$$\sum_{k=0}^{t} S_{k+1} = \sum_{k=0}^{t} S_k + \sum_{k=0}^{t} B_k - \frac{1}{\rho} \sum_{k=0}^{t} I_{k+1}^{(r)}$$

and on the left side of the equation above, shifting the summation index by 1,

$$\sum_{k=1}^{t+1} S_k = \sum_{k=0}^{t} S_k + \sum_{k=0}^{t} B_k - \frac{1}{\rho} \sum_{k=0}^{t} I_{k+1}^{(r)}.$$

Additionally, subtracting $\sum_{k=1}^{t} S_k$ from both sides of the above equation,

$$\sum_{k=1}^{t+1} S_k - \sum_{k=1}^{t} S_k = \sum_{k=0}^{t} S_k - \sum_{k=1}^{t} S_k + \sum_{k=0}^{t} B_k - \frac{1}{\rho} \sum_{k=0}^{t} I_{k+1}^{(r)}$$

and simplifying further

$$S_{t+1} = S_0 + \sum_{k=0}^{t-1} B_k - \frac{1}{\rho} \sum_{k=0}^{t-1} I_{k+1}^{(r)}. \tag{2.2.14}$$

Recall that $S_t = \bar{S} + D_t$, and notice, in particular, that $S_0 = \bar{S} + D_0$ so that equation 2.2.14 becomes

$$\bar{S} + D_{t+1} = \bar{S} + D_0 + \sum_{k=0}^{t-1} B_k - \frac{1}{\rho} \sum_{k=0}^{t-1} I_{k+1}^{(r)}. \tag{2.2.15}$$

Further simplification leads to

$$\sum_{k=0}^{t-1} B_k = \frac{1}{\rho} \sum_{k=0}^{t-1} I_{k+1}^{(r)} + D_{t+1} - D_0. \tag{2.2.16}$$

The parameters in the above equation can be estimated as follows: from Equation 2.2.16, a regression of the cumulative births against the cumulative time series of cases, reveals ρ as the inverse slope of the regression line. Furthermore, the susceptible deviations D_t, correspond to the residuals of the standard linear regression model.

In our particular case, we estimate the transmission parameter β as a constant, from the simulated data, using simple regression. We achieve this by employing the equation of the epidemic process earlier derived in Equation 2.2.9 as

$$\lambda_{t+1} = \beta_s (I_t + \theta_t)^{\alpha_1} (\bar{S} + D_t)^{\alpha_2}.$$

We assume homogeneous mixing in the simulations so that we set $\alpha_1 = \alpha_2 = 1$. The epidemic intensity equation therefore reduces to

$$\lambda_{t+1} = \beta (\bar{S} + D_t)(I_t + \theta_t).$$

We first perform a log-transformation of this equation to obtain

$$log(\lambda_{t+1}) = \log(\beta) + \log(I_t + \theta_t) + \log(\bar{S} + D_t) \tag{2.2.17}$$

In the equation above, I_t and θ_t are known from the data. Additionally, the time series of susceptible deviations D_t, is estimated from Equation 2.2.16, after setting $\rho = 1$ (perfect reporting), as the residuals of the regression of the cumulative births on the cumulative incidences. The mean susceptible size $\bar{S}$, is estimated using marginal likelihood.

We will now proceed to describe how we evaluated the two models.

2.2.5 Assessing the models

Our simulation model was a discrete-time SEIR model having daily time-steps, with infection modelled as a binomial process as described in Section 2.2.2. We assumed the latent and incubation periods to be fixed, with durations of 10 and 6 days respectively.

For the default data sets, we simulated 1500 stochastic epidemics, with the same initial values, for a non-naive population, with demographic processes - births balanced by deaths. Each simulation spanned the period January 1, 2006 to December 31, 2015 but allowed for a burn-in period of 3 years. The simulations were done for four patches. For each simulation run, the population level cases were found by aggregating the cases from the patches.

For each patch, we assumed a constant population size, and constant rate of importation, and transmission. The importation rate was 1 case per 70 days, on average, for all four patches. We also assumed perfect reporting.

The patches were assigned the same population sizes as the 2015 estimates for the four health zones from the Katanga data. The initial parameters for the simulations are summarized in Table 2.1.

Table 2.1: Initial parameter values for the default 1500 simulation runs. The population sizes for the patches are the same as the four health zones from the Katanga province (2015).

Patch	$\mathcal{R}_0$	Population size
1	8.0	308 000
2	10.5	214 000
3	14.2	133 000
4	18.8	285 000

The assessment was done by comparing the values of the initial susceptible population S_0, and the basic reproductive ratio $\mathcal{R}_0$ (TSIR) and $\mathcal{R}_{eff}$ (removal method), used for the simulations, with their distributions obtained after fitting 1500 stochastic epidemics for each of the four patches. Note that when the population is non-naive, the removal method estimates $\mathcal{R}_{eff}$ instead of $\mathcal{R}_0$ because the implicit assumption of the removal method that the total population is the sum of the initial infected and susceptible individuals - that is, the population is naive - is violated. For the TSIR model, we fitted the whole time series (2006-2015). The removal method only works on short time series, hence we fitted only the last year (2015) of the epidemic for each patch and the aggregated data.

Each default simulation run produced an S_0 value after the burn-in period, so that all 1500 simulation runs had a different S_0 value, if the population was assumed to be non-naive. For each simulation, we calculated the ratio of the fitted estimate to the actual. This was done for each patch separately and for the aggregated cases, representing the population estimate. We did not evaluate the $\mathcal{R}_0$ and $\mathcal{R}_{eff}$ estimate at the population level because, in the simulation, the total "reported" cases were simply the sum of cases from the four patches hence, we could not simply assume that the population level transmission parameter was a simple function of the patch-level values.

We modified the simulated data to match the requirements of each simulation method. For the 2015 data sets, some epidemics did not take off as expected - zero cases were produced throughout the course of the epidemic. For those epidemics, we set the fitted estimates to be missing (NA). The number of epidemics that did not take off, out of 1500, for patches 1, 2, 3, and 4 were 420, 506, 581 and 608, respectively. For the epidemics

that took off, we truncated each data set by removing all leading zeros. We replaced all spontaneous zero cases in the remaining data with single cases. The resulting time series were then used for the removal fitting and assessments.

For the TSIR data fitting, we replaced all zero cases with an arbitrarily chosen value close to zero to resolve an issue resulting from the log-transformation applied to Equation 2.2.17. Recall that the resulting equation after the transformation is $\log(\lambda_{t+1}) = \log(\beta) + \log(I_t + \theta_t) + \log(\bar{S} + D_t)$, which evaluates to $-\infty$ when the cases at any time are zero. Replacing the 0's with 0.00001 resolved this issue. Since the simulated data was generated assuming perfect reporting, we skipped the susceptible reconstruction part of the TSIR fitting procedure because we could calculate the exact proportion of susceptible individuals at each time step. One of the assumptions of the TSIR model is that the transmission parameter varies seasonally but we assumed a constant transmission parameter for both the simulation and the fitting process.

After running the assessments on the default data sets which assumed the population was non-naive with demographic processes, the TSIR model fits were desirable but the removal method produced highly biased estimates. In order to tease out what was driving the biases in the removal estimates, we simulated two more sets of 1500 stochastic realizations representing intermediate assumptions between the default simulation model and the assumptions of the chain binomial model that underlies the removal method, specifically. In one simulation, we assumed the population to be naive and not influenced by births and deaths, and in the other, we assumed the population to be naive and influenced by births that were balanced, on average, by deaths. The former set of simulation assumptions were the closest to those of the removal method. Most of the other parameters - population size, immigration rate, duration of infectiousness, and so forth - remained the same as the default simulations. Using these data sets, we estimated S_0 and the basic reproductive ratio, $\mathcal{R}_0$ for comparison with the true values used in the simulations. Note here that the removal assumption that the total population is the sum of the initial number of susceptible and infected individuals was implemented in the simulations so that we were estimating $\mathcal{R}_0$, not $\mathcal{R}_{eff}$ as in the default simulation.

The results of the assessments are presented in Section 2.3 which follows.

2.3 Results

2.3.1 Removal method estimates

2.3.1.1 Non-naive population, with births balanced by deaths (default simulation)

In this section, we present the results of fitting the default epidemic simulations - assuming the population was non-naive, with births balanced by deaths - with the removal method to obtain estimates of $\mathcal{R}_{eff}$ for the patches, and S_0 for each patch and the aggregated data.

In order to estimate $\mathcal{R}_{eff}$, we aggregated the data into biweekly bins and replaced all zero cases with 1, after we had truncated all the zero cases leading up to the first non-zero case. Recall we mentioned that some epidemics did not take off (resulted in zero cases throughout); we set the removal fit of those epidemics to NA.

Figure 2.1 presents the distribution of the ratio of estimated to actual $\mathcal{R}_{eff}$ estimates for the four patches. All the $\mathcal{R}_{eff}$ estimates were over-estimated.

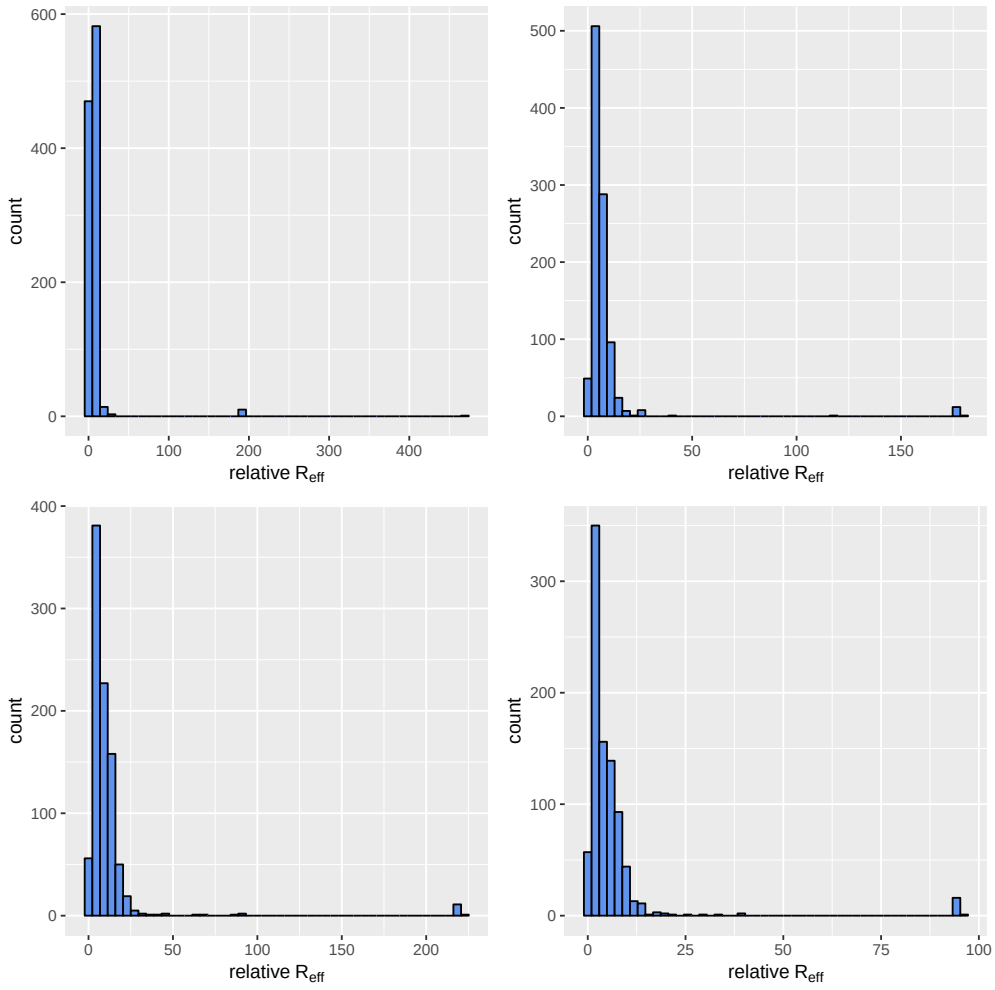


Figure 2.1: Ratio of estimated to actual $\mathcal{R}_{eff}$ values for four patches, using the removal method. The simulation assumes each patch has a non-naive population with a constant birth rate balanced out by the death rate - this was the default simulation - but the removal method assumes the population to be naive, without births and deaths. The assumptions of the data generation process and the removal method do not match in this case. The number of epidemics that did not take off for patches 1, 2, 3, and 4 were 420, 506, 581 and 608, respectively. The plots are arranged from top left to bottom right according to patch 1, patch 2, patch 3, and patch 4, respectively. The reference point is at 1 on the x-axis. Notice that the axes are not on the same scale.

Table 2.2 below shows a summary of the results of the ratio of estimated to actual $\mathcal{R}_{eff}$ values across the four patches in terms of mean and median.

Table 2.2: Median and mean of the ratio of the estimated to actual $\mathcal{R}_{eff}$ for the four patches, using the removal method. The simulation assumes the population is non-naive, with births balanced by deaths. All the $\mathcal{R}_{eff}$ estimates were biased high.

	Patch 1	Patch 2	Patch 3	Patch 4
Median	5.12	5.00	7.12	3.46
Mean	7.87	8.31	11.64	6.17

In Figure 2.2 which follows, we show the distribution of the ratio of the estimated to actual S_0 values for the four patches. It is clear that S_0 was over-estimated in all patches.

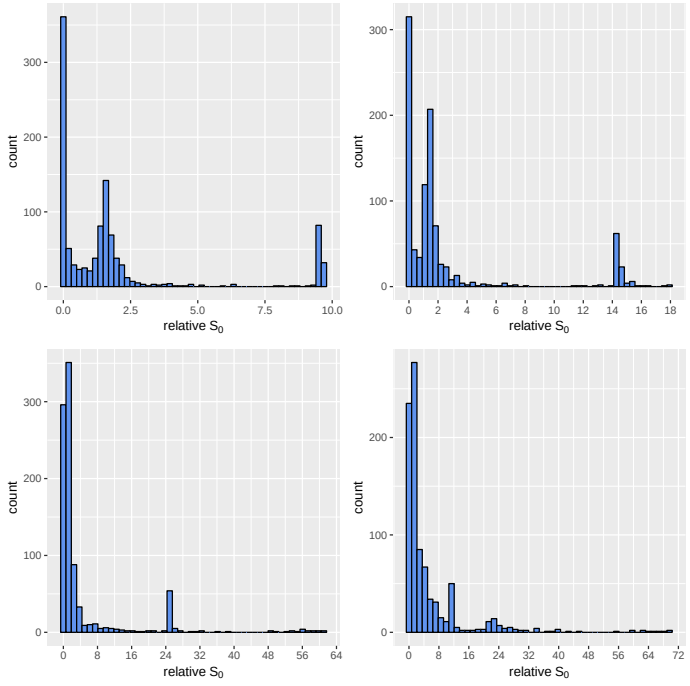


Figure 2.2: Ratio of estimated to actual S_0 values, for the four patches, using the removal method. The simulation assumes that each patch has a non-naive population with constant births, balanced by deaths, contrary to the assumptions of the removal method that the population is naive, without births and deaths. The number of epidemics that did not take off for patches 1, 2, 3, and 4 were 420, 506, 581 and 608, respectively. The plots are arranged from top left to bottom right according to patch 1, patch 2, patch 3, and patch 4, respectively. The reference point is at 1 on the x-axis. Notice that the axes are not on the same scale.

Figure 2.3 illustrates that S_0 is also over-estimated in the aggregated data.

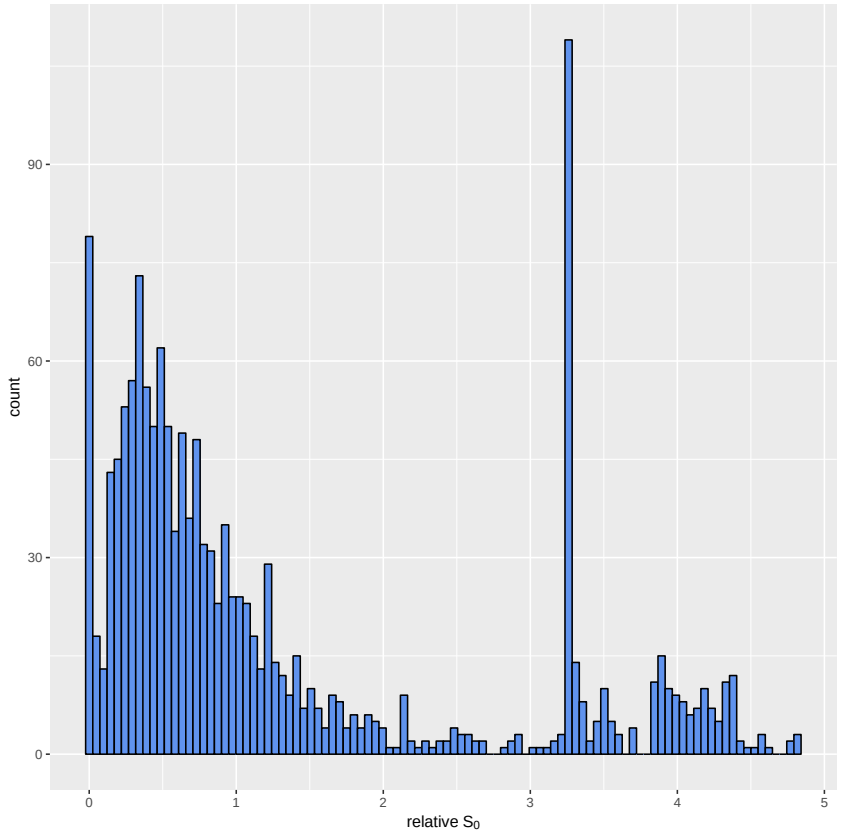


Figure 2.3: Ratio of estimated to actual S_0 values, for the aggregated cases, using the removal method. The simulation assumes that the population is non-naive, with a constant birth rate, balanced out by deaths, on average. This assumption is a mismatch to the removal method's assumption of a naive population without births and deaths. Twenty five epidemics did not take off in the aggregated epidemics. Note that the reference point is at 1.

A summary, in terms of mean and median, of the results of the ratio of the estimated to actual S_0 values, across the four patches and the aggregated data, are presented in Table 2.3 which follows.

Table 2.3: Median and mean of the ratio of the estimated to actual S_0 for the four patches. The simulation assumes the population is non-naive, with births balanced by deaths. According to the median, all the S_0 estimates were biased high at the patch level and low at the aggregated level.

	Patch 1	Patch 2	Patch 3	Patch 4	Aggregated
Median	1.26	1.26	1.23	1.42	0.73
Mean	1.90	2.51	4.64	5.18	1.28

From the results of this section, we realized that the S_0 and $\mathcal{R}_{eff}$ estimates across all four patches, and the aggregated data, were biased due to the mismatching assumption between the data generation process and the removal method. Hence, to tease out what was driving the biases, we simulated 1500 stochastic epidemics, assuming the population was naive with demographics (births balanced by deaths), and fitted them with the removal method. The results are presented in the next section.

2.3.1.2 Naive population, with demographics (births balanced by deaths)

Here, the simulated data was modified as in Section 2.3.1.1, to put it in a from usable by the removal method. We proceed to present the results.

Figure 2.4 represents the relative - estimated to actual - $\mathcal{R}_0$ estimates for the four patches assuming the simulated population is naive, and with demographic processes (births balanced by deaths).

As can be seen in the figure, when births and deaths influenced the dynamics of the naive population, the removal method over-estimated $\mathcal{R}_0$ in the first two patches and under-estimated $\mathcal{R}_0$ in the last two.

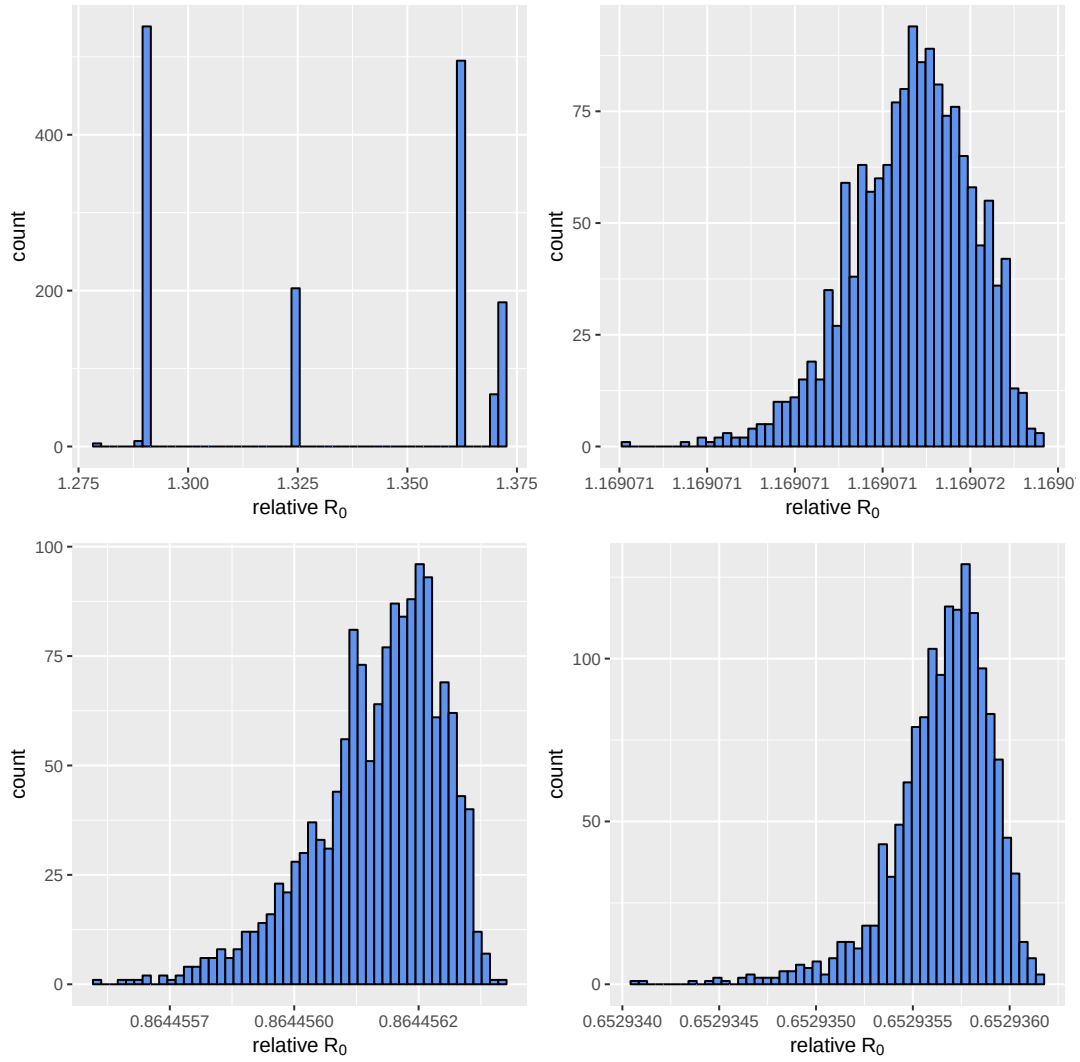


Figure 2.4: Ratio of estimated to actual $\mathcal{R}_0$ values for the four patches, using the removal method. The simulation assumes each patch has a naive population, with demographic processes, but the removal method assumes that the population is naive without births and deaths. Hence, there is a mismatch in the assumptions about births and deaths. The plots are arranged from top left to bottom right according to patch 1, patch 2, patch 3, and patch 4, respectively. Note that the reference point is at 1 on the x-axis. Notice, also, that the axes are not on the same scale.

In what follows, Table 2.4 contains a summary of the results displayed above across all four patches, in terms of mean and median.

According to the medians, it is clear that $\mathcal{R}_0$ was biased high in patches 1 and 2 and biased low in patches 3 and 4.

Table 2.4: Median and mean of the ratio of the estimated to actual $\mathcal{R}_0$ for the four patches. The simulation assumes the population is naive, with births balanced by deaths. All the $\mathcal{R}_0$ estimates were biased high.

	Patch 1	Patch 2	Patch 3	Patch 4
Median	1.36	1.17	0.86	0.65
Mean	1.34	1.17	0.86	0.65

Figure 2.5 shows that S_0 was over-estimated in all patches. Observe, however, that the fitted S_0 estimates, even though over-estimated, were proximate to the actual.

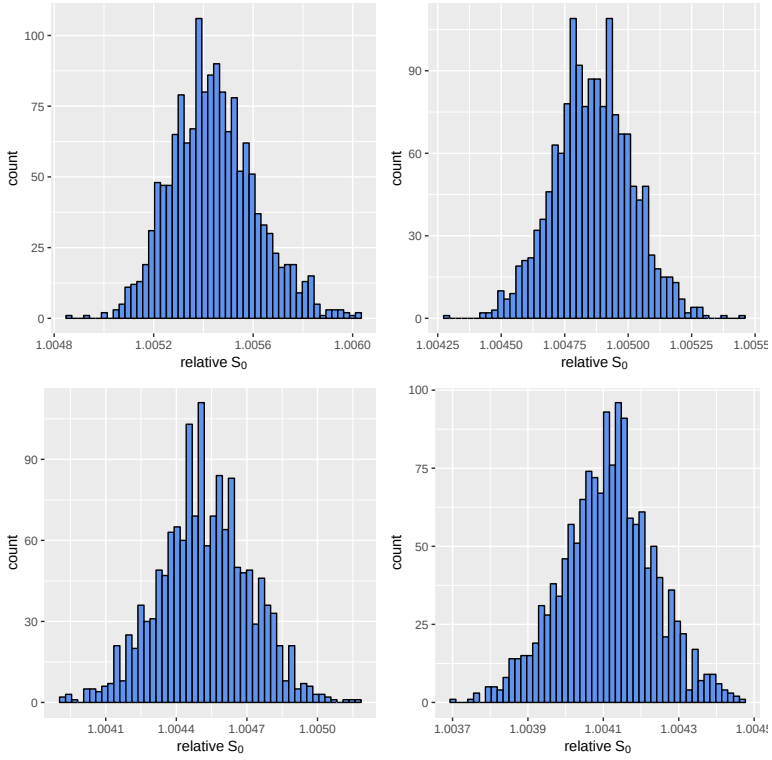


Figure 2.5: Ratio of estimated to actual S_0 values, for four patches, using the removal method The simulation assumes each patch has a naive population with a constant birth and death rate but the removal method assumes the population is naive without births and deaths. Hence, there is a mismatching assumption about births and deaths. The plots are arranged from top left to bottom right according to patch 1, patch 2, patch 3, and patch 4, respectively. The reference point is at 1 on the x-axis. Notice that the axes are not on the same scale.

Figure 2.6 shows that the S_0 was under-estimated in the aggregated data.

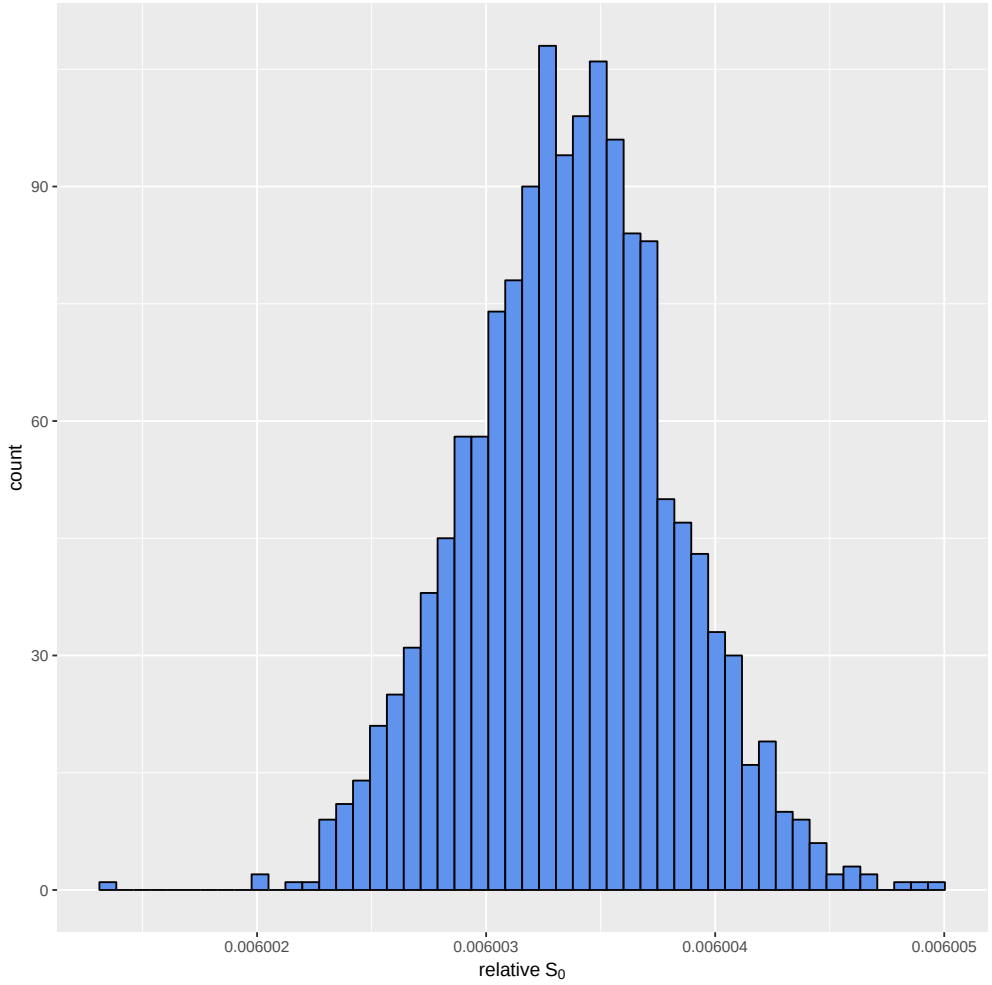


Figure 2.6: Ratio of estimated to actual S_0, for the aggregated cases, using the removal method. The simulation assumes the population is naive, with demographics, but the removal method assumes the population is naive without demographics, hence a mismatch in assumptions about demographics. Note that the reference point is at 1 on the x-axis.

A summary of the results of the ratio of the estimated to actual S_0 values, in terms of mean and median across the four patches, and the aggregated data, are presented in Table 2.5, which follows.

Table 2.5: Median and mean of the ratio of the estimated to actual S_0 for the four patches, the aggregated data. The simulation assumes the population is naive, with demographics. All the S_0 estimates were unbiased at the patch level, but biased low at the aggregated level.

	Patch 1	Patch 2	Patch 3	Patch 4	Aggregated
Median	1.01	1.00	1.00	1.00	0.01
Mean	1.01	1.00	1.00	1.00	0.01

To investigate, the biases further, we simulated 1500 stochastic epidemics with assumptions closely matching that of the removal method, that is, the population is naive, without demographic processes. The results are presented in the following section.

2.3.1.3 Naive population, without births and deaths

Here also, the data was modified using the method in Section 2.3.1.1, and the S_0 and $\mathcal{R}_0$ estimates were analysed. The results follow.

Figure 2.7 shows the relative $\mathcal{R}_0$ estimates for the four patches, when the simulation assumed the populations were naive, without births and deaths, closely matching the removal method assumptions.

From the figure, we can see $\mathcal{R}_0$ was over-estimated in patches 1 and 2, and underestimated in patches 3 and 4. The results here closely resemble those of the previous section (Figure 2.4) where the simulation assumed the population was naive, and with demographic processes.

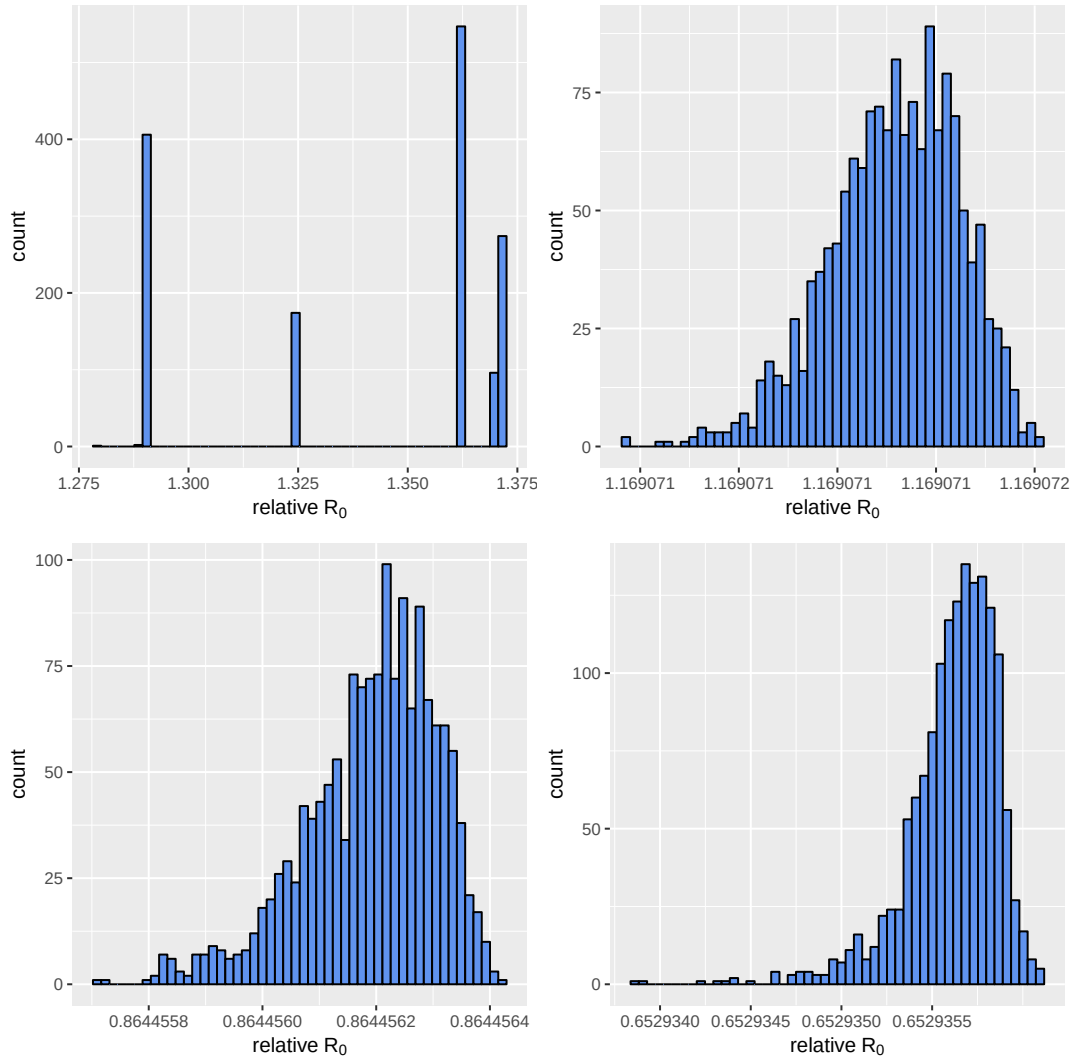


Figure 2.7: Ratio of estimated to actual $\mathcal{R}_0$ values for the four patches, using the removal method. The simulation assumes each patch population is naive, without births and deaths. The simulation assumptions here are similar to those of removal method. The plots are arranged from top left to bottom right according to patch 1, patch 2, patch 3, and patch 4, respectively. The reference point is at 1 on the x-axis. Notice that the axes are not on the same scale.

In Table 2.6 below, we present a summary of the results, in terms of mean and median across the four patches, of the ratio of the estimated to actual $\mathcal{R}_0$ values when the simulation assumes the population is naive without demographics, similar to the assumptions of the removal method.

Table 2.6: Median and mean of the ratio of the estimated to actual $\mathcal{R}_0$ for the four patches. The simulation assumes the population is naive, without births and deaths, similar to the assumptions of the removal method. The $\mathcal{R}_0$ estimates were biased high in patches 1 and 2, and biased low in patches 3 and 4.

	Patch 1	Patch 2	Patch 3	Patch 4
Median	1.33	1.17	0.86	0.65
Mean	1.33	1.17	0.86	0.65

Figure 2.8, illustrates the relative S_0 estimates at the patch level. The estimates are unbiased since their relative ratios surround 1.

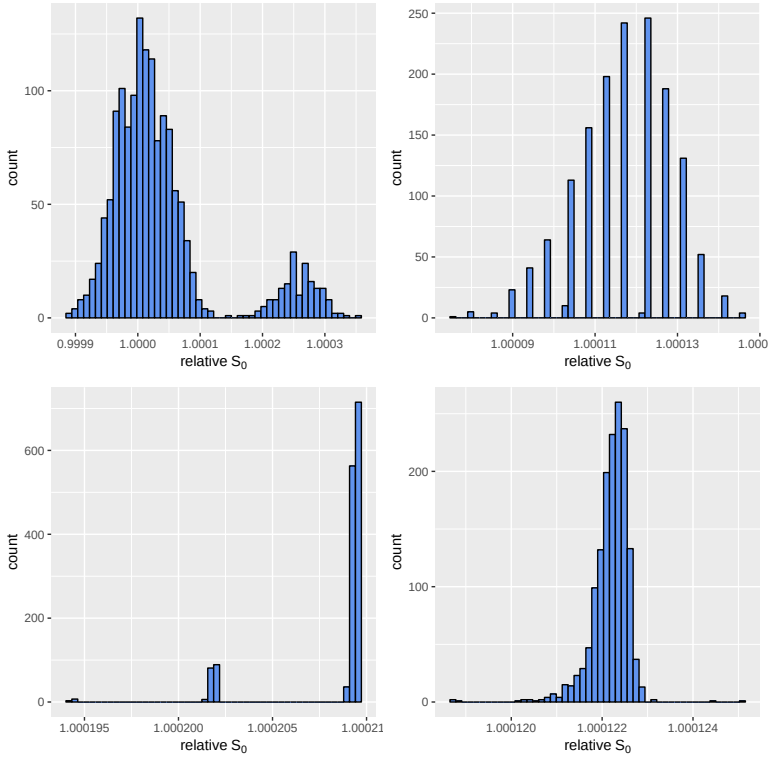


Figure 2.8: Ratio of estimated to actual S_0 values, for the four patches, using the removal method. The simulation assumes that each patch has a naive population without demographics. These assumptions are the closest to that of the removal method. The plots are arranged from top left to bottom right according to patch 1, patch 2, patch 3, and patch 4, respectively. The reference point is at 1 on the x-axis. Notice that the axes are not on the same scale.

Figure 2.9 demonstrates that S_0 at the aggregated level was under-estimated when the simulation assumed that the population is naive, without demographics, similar to the assumptions of the removal method.

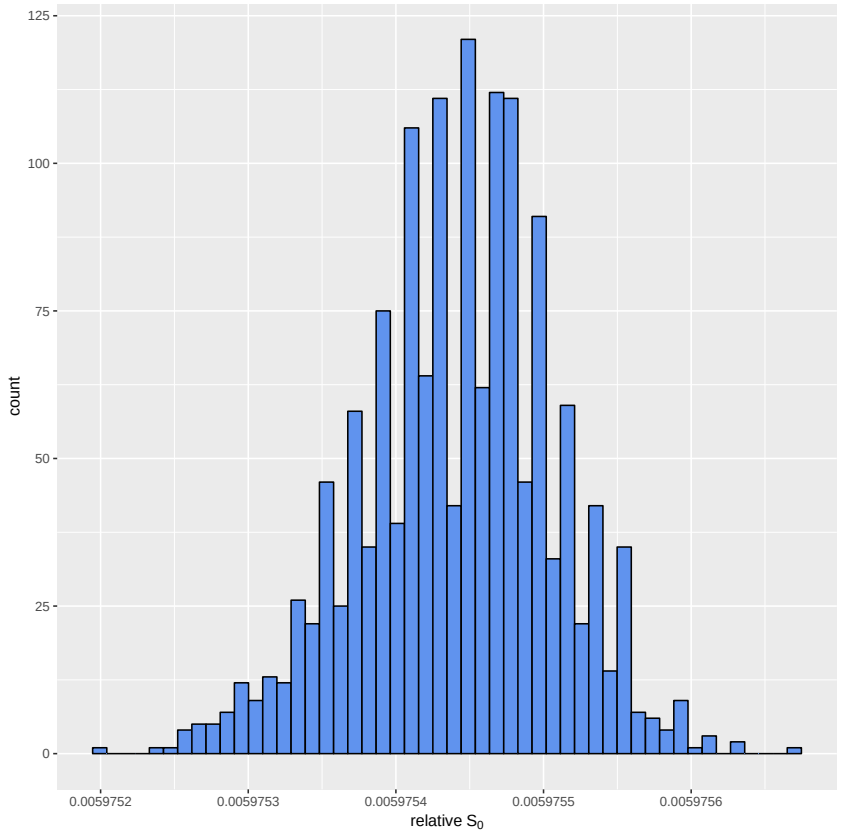


Figure 2.9: Ratio of estimated to actual S_0, for the aggregated cases, using the removal method. The simulation assumes the population is naive, and without births and deaths, therefore matching the assumptions of the removal method. Note that the reference point is at 1 on the x-axis.

In Table 2.7 which follows, these results of the ratio of the estimated to actual S_0 values above are summarised in terms of the mean and median across the four patches, and the aggregated data.

Table 2.7: Median and mean of the ratio of the estimated to actual S_0 for the four patches, and the aggregated data. The simulation assumes the population is naive, without demographics. This assumption closely matches that of the removal method. The S_0 estimates for the patches were unbiased for the patches, but biased low for the aggregated data.

	Patch 1	Patch 2	Patch 3	Patch 4	Aggregated
Median	1.00	1.00	1.00	1.00	0.01
Mean	1.00	1.00	1.00	1.00	0.01

Table 2.8 which follows, sums up all the results of the removal method fits, and the biases realized for the three simulation assumptions.

Table 2.8: Summary of the removal estimation results, with the biases that resulted, and the figures that depict them. The simulation assumptions are also included.

	Simulation assumptions		Estimation summary		Figure number
	Initial population	Demography	Quantity	Result	
Default	Non-naive	Yes	$\mathcal{R}_{eff}$ (patch)	biased (high)	2.1
			S_0 (patch)	biased (high)	2.2
			S_0 (aggregated)	biased (low)	2.3
Intermediate	Naive	Yes	$\mathcal{R}_0$ (patch)	biased (mixed)	2.4
			S_0 (patch)	unbiased	2.5
			S_0 (aggregated)	biased (low)	2.6
Matched	Naive	No	$\mathcal{R}_0$ (patch)	biased (mixed)	2.7
			S_0 (patch)	unbiased	2.8
			S_0 (aggregated)	biased (low)	2.9

2.3.2 TSIR estimates

In this section, we present the results of fitting the default 1500 stochastic epidemics, each spanning the period between December 26, 2005 and December 28, 2015, with the TSIR model. Recall that for the default simulation, we assumed that the population was non-naive and that the population dynamics were influenced by births, which were balanced out by deaths, on average.

Because we assumed perfect reporting in our simulations, the actual S_0 value at the patch level, were summed to yield the population-level S_0 value to allow for comparison with the fitted value. The R_0 estimates were only fitted at the patch-level because the population-level R_0 was not assumed to be the sum of R_0 from the patches. The TSIR model was only fitted to the default simulated epidemics because the simulation assumptions were similar to that of the TSIR model.

In the following section, we display the results of comparing the fitted S_0 and $\mathcal{R}_0$ estimates to the actuals.

2.3.2.1 Non-naive population, with demographics (default simulation)

To estimate $\mathcal{R}_0$ from the default simulation data, we replaced all zero cases with a value of 0.00001, then we aggregated the data into biweekly bins to match the time-step of the TSIR model.

From Figure 2.10, we observe in all four patches that $\mathcal{R}_0$ was biased low, since most of the estimates lie to the left of the reference of 1 on the x-axis.

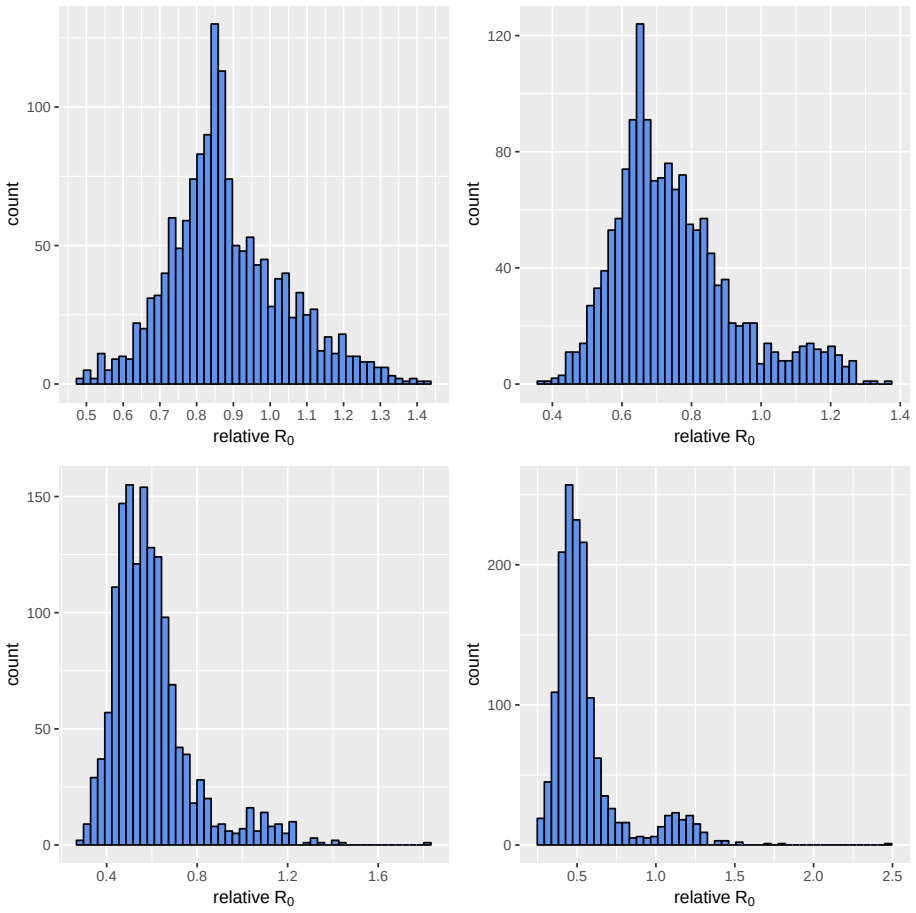


Figure 2.10: Ratio of estimated to actual $\mathcal{R}_0$ values for each of the four patches, using the TSIR model. The simulation assumes the population is non-naive, with demographic processes. This closely matches some of the the assumptions of the TSIR model. The plots are arranged from top left to bottom right according to patch 1, patch 2, patch 3, and patch 4, respectively. The reference point is at 1 on the x-axis. Notice that the axes are not on the same scale.

Table 2.9 summarizes the above results, in terms of mean and median across the four patches, of the ratio of the estimated to actual $\mathcal{R}_0$ values. It is clear that $\mathcal{R}_0$ was under-estimated across all patches and hence, biased low.

Table 2.9: Median and mean of the ratio of the estimated to actual $\mathcal{R}_0$ for the four patches, using the TSIR model. This is the default simulation, which assumes the population is non-naive, with demographic processes. The assumptions of the default simulation closely match that of the TSIR. All the $\mathcal{R}_0$ estimates were biased low.

	Patch 1	Patch 2	Patch 3	Patch 4
Median	0.86	0.72	0.57	0.49
Mean	0.88	0.75	0.60	0.58

Figure 2.11 presents the ratio of the estimated to actual S_0 estimate across all four patches for the default simulation data - assuming a non-naive population with demographics. It is clear that the S_0 estimates were mostly under-estimated for patches 1 and 2, and mostly over-estimated for patches 3 and 4.

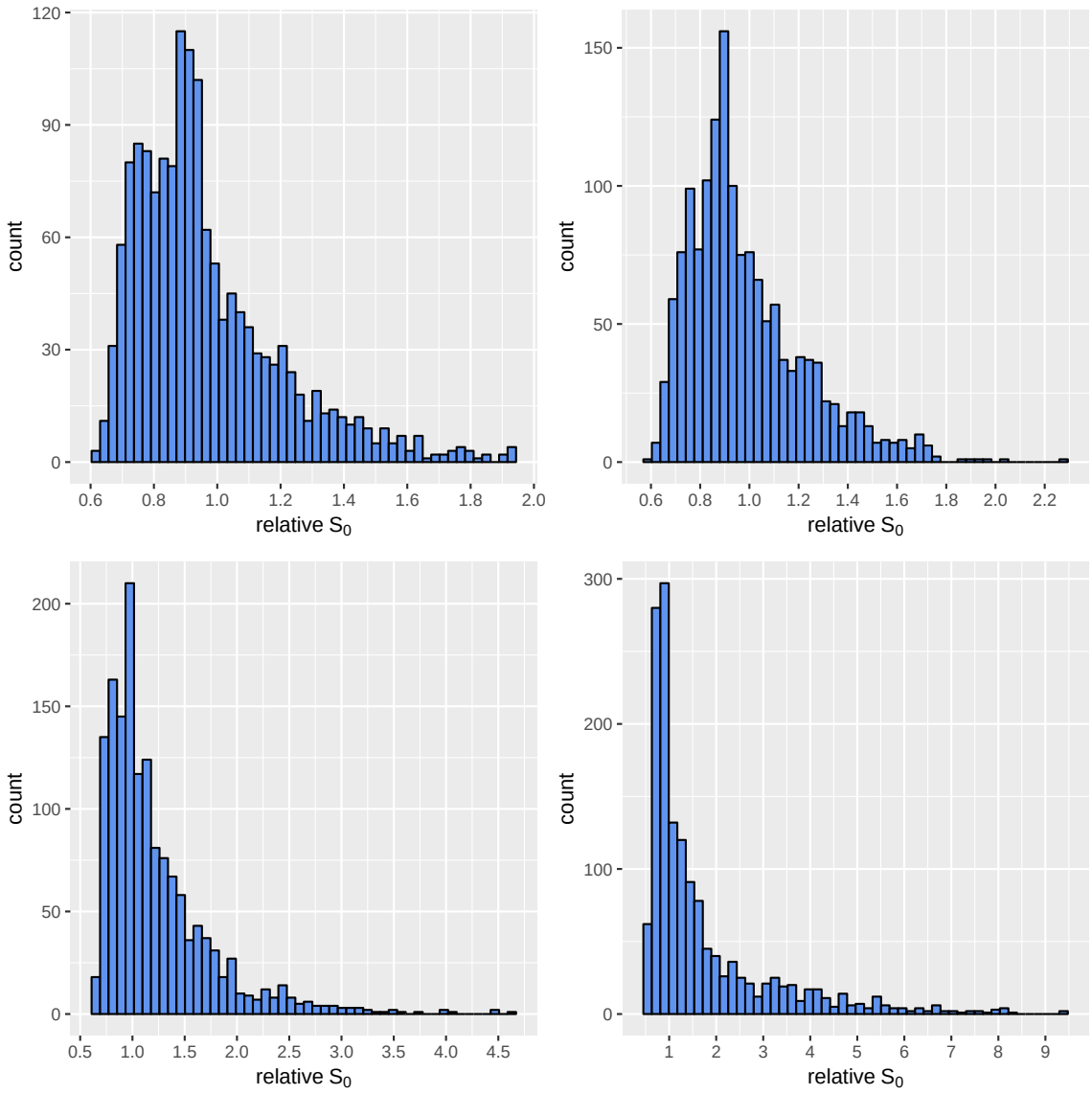


Figure 2.11: Ratio of estimated to actual S_0 values, for each of the four patches, using the TSIR model. The simulation assumes a non-naive population with demographics. This closely matches some of the assumptions of the TSIR model. The plots are arranged from top left to bottom right according to patch 1, patch 2, patch 3, and patch 4. The reference point is on 1 on the x-axis. Notice that the axes are not on the same scale.

Figure 2.12 shows the results of the relative - estimated to actual - S_0 estimates of the aggregated data.

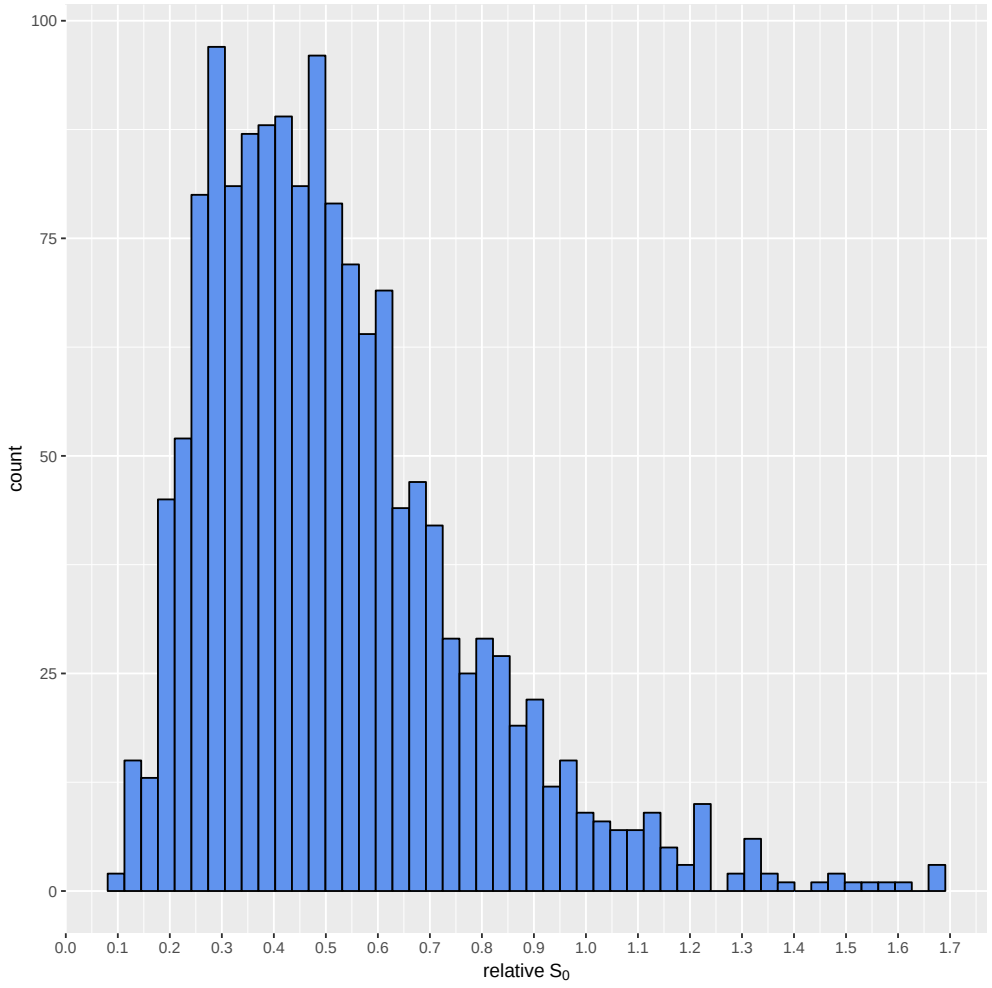


Figure 2.12: Ratio of estimated to actual S_0 values of the aggregated cases, using the TSIR method. The simulation assumes a non-naive population with demographics. The reference point is at 1 on the x-axis.

In Table 2.10 which follows, we summarise the results in Figures 2.11 and 2.12, in terms of mean and median across the patches and the aggregated data.

Table 2.10: Median and mean of the ratio of the estimated to actual S_0 for the four patches and the aggregated data. The S_0 estimates for patches 1 and 2 were biased low, whiles those of patch 4 were biased high. The patch 3 estimates were nearly unbiased, and the aggregated data estimates were biased low.

	Patch 1	Patch 2	Patch 3	Patch 4	Aggregated
Median	0.91	0.92	1.07	1.15	0.47
Mean	0.97	0.98	1.23	1.72	0.52

In Table 2.11 which follows, we have summarized the results of assessing the TSIR model, after it was fitted with the default simulation data, and the biases that were realized.

Table 2.11: Summary of the TSIR results, with the biases that resulted, and the figures that depict them. The TSIR was used to fit only the default simulation data, which assumed that the population was non-naive, with demographic processes.

	Simulation assumptions		Estimation summary		Figure number
	Initial population	Demography	Quantity	Result	
Default	Non-naive	Yes	$\mathcal{R}_0$ (patch)	biased (low)	2.10
			S_0 (patch)	biased (mixed)	2.11
			S_0 (aggregated)	biased (low)	2.12

2.4 Discussion

We have demonstrated the importance of understanding the underlying assumptions of the TSIR model, and the removal method, prior to using them to fit measles epidemic data. We find that when important fundamental processes - births, deaths, and immunity - underlying how the epidemic data are generated, match the assumptions of the fitting model, reasonable estimates can be obtained. In essence, the observed data should conform to certain aspects of the fitting model's assumptions in order to achieve

reasonable estimates. It is interesting to know that the fitting is robust to some differences in assumptions, especially in terms of the duration of infection, and the model time-step.

This result was facilitated by running stochastic epidemic simulations, with varying assumptions about the occurrence of births and deaths, and whether the population was naive or not. This study only confirms the utility of simulations as a powerful tool for pursuing scientific enquiries. Simulations are a useful tool for investigating inquiries surrounding various phenomena. Refer to Bellan *et al.* (2015), Halloran *et al.* (2017), and Lessler *et al.* (2016) for examples of cases in which simulations were employed to answer pertinent questions in mathematical epidemiology.

The amount of bias surrounding both the TSIR and removal S_0 estimates at the aggregated level, compared to the patch level, is striking. This seems to suggest that parameter fits on data at an aggregated scale tend to have higher amounts of uncertainty compared to data reported at a finer spatial scale. Hence, statistical inference done with fitted estimates of aggregated data should be interpreted with cognisance of the potential biases resulting from aggregation. Even though we assumed perfect reporting in our simulations, the aggregated data produced more biased S_0 estimates than the segregated patch data. It is, therefore, interesting to imagine the amount of bias that could result from actual epidemics, knowing they are often aggregated and also affected by issues such as under-reporting (Bobashev *et al.*, 2000).

Several reasons could explain the biases surrounding the removal estimates. The biases are driven by whether the population being fitted is naive or not - estimates from the non-naive populations are highly biased compared to those from the naive populations, irrespective of assumptions about the occurrence of demographic processes. In their paper, Ferrari *et al.* (2005) indicated that the removal estimator's performance is affected by the depleting susceptible population along the course of the epidemic - the method assumes there is no susceptible recruitment through birth, due to the fast progression of the epidemic. Hence, applying the model to the full course of a long epidemic could be problematic. This could be one reason behind the biases observed since we applied the method to the 2015 time series, beginning from the first occurrence of non-zero cases on the epidemic trajectory. Additionally, the removal method does not work on data that has spontaneous zeros. This assumption is difficult to adhere to, especially considering the fact that when the observed data is segregated at a fine spatial scale, zero cases are prevalent. When we replace all zero cases with 1's to resolve this issue, it potentially biased the expected $\mathcal{R}_0$ as depicted in Figure 2.7. However, the S_0 estimates at the patch level tend to be unaffected by this fix since they remain unbiased, but are biased low at

the aggregated level, as can be seen in Figures 2.9 and 2.8. The biases in the fitted S_0 estimates are very little, compared with those of $\mathcal{R}_0$ and $\mathcal{R}_{eff}$.

Some of the removal estimates were occasionally large, relative to the cluster of estimates. Our preliminary investigation reveals that these outliers occurred when the epidemic had a single peak, or when the epidemic did not take off. We will be investigating this further in the future.

The approach of replacing zero cases with 1's is barely one out of many potential "fixes". Other approaches could include any attempt at making sure all cases are non-zero, including "zooming" in to the part of the epidemic with a continuous succession of non-zero cases. In implementing such "fixes", attempts should be made to quantify the potential consequential biases.

The TSIR model estimates of S_0 are distributed around the reference point of 1, so that at the patch level, they are a mix of biases - patches 1 and 2 are biased low, patch 3 is unbiased, and patch 4 is biased high, but all are close to 1 - but at the aggregated level, they are biased low. The $\mathcal{R}_0$ estimates are biased low for all the patches. These biases could have been introduced when we replaced zero cases with a value close to zero. This "fix", however, had a marginal magnitude of bias on S_0 than it did on $\mathcal{R}_0$ - since it only partially increased the number of cases in the data, as confirmed by the results. One way to avoid this fix is to employ a Bayesian framework in the parametrizing the TSIR (Ferrari *et al.*, 2008), instead of the linear regression method which employs the log-transformation, disallowing for zero cases in the data.

When some of the assumptions underlying the data generation process closely matched the assumptions of the estimation model, the two models predicted S_0 better at the patch level than at the population level, with the removal S_0 estimate being proximate to the actual simulation values - for the naive populations. The TSIR predicted low biased estimates of $\mathcal{R}_0$, whiles the removal method over-estimated (produced high biased) $\mathcal{R}_{eff}$ estimates for the non-naive populations. When we teased out what was driving the bias in the removal method, it became clear that the assumption of naiveness in the population was a major driver - the method produced a mix of biased estimates of $\mathcal{R}_0$ and S_0 for the naive populations, which were relatively better - to an order of magnitude - than the estimates obtained from the non-naive populations.

The results of this study should be interpreted with caution. It is our belief that each different artificial modification of the time series that may be necessary to use an estimation method may produce different biases in magnitude and direction. Also, since the processes may differ significantly between those underlying how the actual data is generated and the simulated data, this exercise could be more complicated than what

we have presented here.

We have provided insights into the performance of the TSIR model in terms of the biases produced, when a non-naive population, influenced by demographic processes, is fitted with the model to obtain estimates of $\mathcal{R}_0$ and S_0. We have also provided the same information regarding the removal fits of $\mathcal{R}_{eff}$ or $\mathcal{R}_0$, and S_0 from three different kinds of populations: non-naive, with births, balanced by deaths; naive, with births, balanced by deaths; naive, without births and deaths. It is evident that the performance of the TSIR model and removal method depends on which important assumptions are matched between the estimation model and the data, the degree of aggregation of the data, the artificial modifications made to the data to put it in a format usable by the estimation method, and the parameters being estimated.

In summary, the insights to be drawn from this chapter are that: violations of certain core assumptions of the estimation model - for instance, the use of the removal method to fit data from non-naive populations with high birth and death rates - can lead to biased estimates being produced; parameter estimates appear fairly robust to simple modifications of the time series, such as, replacing zero cases with 1s, for use with the removal method, or replacing zero cases with small numbers, for use with the TSIR method. As to whether these findings hold when data are imperfectly reported remains to be investigated.

Chapter 3

Measles outbreak in a high school in South Africa: assessing the impact of a reactive vaccination campaign

3.1 Introduction

In February 2017 a measles outbreak was declared in Stellenbosch, South Africa, after five cases were detected in an all-boys high school. According to the South African Department of Health, the occurrence of five laboratory-confirmed measles cases warrants the declaration of an outbreak in the affected area. A day after the outbreak was declared, a non-selective reactive vaccination campaign was kick-started and more than 93% of students within the school were vaccinated in a day.

As earlier mentioned in Chapter 1, assessing the impact of a vaccination campaign is essential for improving future campaigns. One common method of appraising a vaccination campaign is to quantify the cases averted. Accordingly, following the vaccination campaign at the high school, our study seeks to assess its impact using the same metric.

Also, whiles the outbreak was ongoing, it was in the interest of the stakeholders - the school authorities, the learners, and the learners' parents - to know how many more cases were expected, and what time remained for the epidemic to subside following the campaign. This was important for two reasons: (1) the stakeholders needed some reassurance, and (2) plans needed to be made towards resuming the school's activities which were put on hold because of the outbreak. This study, therefore, aimed at providing information pertaining to the epidemic progression in terms of cases and time remaining until the epidemic ended.

In the past, dynamical models have been a useful tool for estimating the aforementioned quantities to assess the impact of vaccination programs - Examples of such studies are Ferrari *et al.* (2014); Grais *et al.* (2007); Pinner *et al.* (1992) - hence, we developed and used a dynamical model for closed populations to achieve the set objectives.

3.2 Materials and methods

This section describes the model and methods employed to simulate the measles epidemic as well as the data employed for the model validation.

3.2.1 Data description

3.2.1.1 A measles outbreak in a school in Stellenbosch, South Africa.

On January 15, 2017, Paul Roos Gymnasium, an all-boys high school in Stellenbosch, South Africa, recorded a case of measles. Following the first case, five secondary cases were reported by January 31. This prompted the declaration of an outbreak on February 2, by the Department of Health in the Western Cape Province. As a consequence, a vaccination campaign was held on February 3, 2017 and more than 93% of the 1200 learners at the school were vaccinated. By February 6, the number of lab-confirmed cases among the learners had risen to a total of 23. Subsequently, no new cases were reported. The epidemic in the school, therefore, lasted for about 22 days.

3.2.1.2 Outbreaks in other schools

To validate the model we formulated, we used comparable data sets obtained from published epidemics from five other schools. A description of each epidemic is given below.

The first is an outbreak in School A, Oxford, in 1929. The total number of students was 56, of which there were 36 susceptible students. The remaining 20 students had had measles previously. Since there was no measles vaccine then, all the susceptible students were eventually infected by two students who got infected after illegally visiting the school kitchen and contracting the infection from the cook. The outbreak lasted for 47 days (Hobson, 1934).

In 1930, an outbreak occurred in School B, Oxford. This school had 121 students, of which 68 were susceptible. Out of this number, 48 acquired the infection and 20 remained uninfected. The total epidemic duration was 49 days (Hobson, 1934).

Another outbreak occurred in School B, Oxford, in 1934. The total students were 99 with 48 having previously been infected. The remaining 51 students were therefore

susceptible. There were 11 cases reported during that outbreak. It took 64 days for the epidemic to end (Hobson, 1934).

School A, Oxford experienced another outbreak in 1938. Out of the 95 students, 56 were susceptible and all of them acquired the infection. The epidemic lasted 46 days (Hobson, 1938).

The last epidemic data is of a boarding school outbreak in Pennsylvania, U.S, in 2003. The student population was 663 in total. There were 9 laboratory-confirmed cases, 8 of whom were students and the remaining 1, a staff member. Of all the students, 655 had proper records of previous vaccination uptake (Yeung *et al.*, 2005). The epidemic lasted 39 days.

Apart from the 2003 epidemic, the others occurred in an era where there was no measles vaccine.

3.2.2 Model formulation and calibration

We developed a discrete-time susceptible-exposed-infected-recovered (SEIR) compartmental simulation model. The model was stochastic in nature, and the time-steps were in days. At any time t, the number of individuals in one of four disease states was calculated, that is, susceptible $S(t)$, exposed (latent) but not yet infectious $E(t)$, infectious $I(t)$, and recovered (immune) $R(t)$. The latent $E(t)$, and infectious $I(t)$ groups were further divided into ten and six sub-classes, respectively (refer to the assumptions in Section 3.2.2.1). Depending on the number of days an individual was exposed or infectious, they were classified into one of: $E_1(t), E_2(t), \ldots, E_{10}(t)$ and $I_1(t), I_2(t), \ldots, I_6(t)$. That is, an individual classified as $E_n(t)$ was infected at time $t-n-1$, whereas an individual classified as $I_n(t)$ was infected at time $t-n-11$.

In the model, new infections are produced at the next time step $t+1$ with a binomial probability, which has as parameters, the number of susceptible individuals at time t, and a probability which depends on the total number of infectious individuals in the population at that time.

Mathematically, this is written as

$$E_1(t+1) \sim \text{Binom}(S(t), p(t))$$

where

$$p(t) = 1 - \exp\left(-\frac{\beta}{N}\sum_{t=1}^{6} I(t)\right). \tag{3.2.1}$$

Immunization was modelled as a one-time event, reflecting the nature of the campaign. We assigned this event a binomial distribution with the number of susceptible

individuals at the time of the campaign and the observed vaccinated proportion in the high school's campaign $p_v(t)$ as input parameters.

Figure 3.1 illustrates the structure of the model we have described.

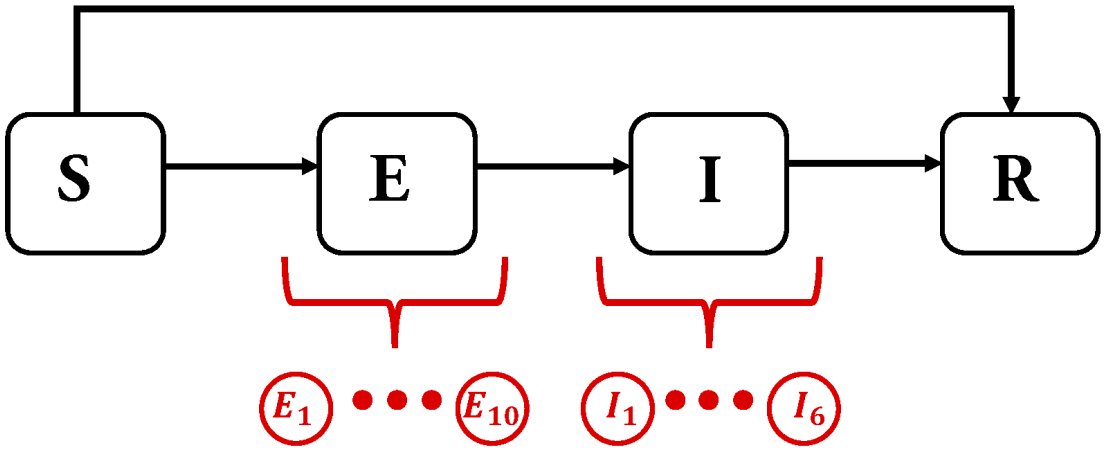


Figure 3.1: Basic diagram of the measles SEIR model. Susceptible individuals acquire the infection and move to the exposed class for ten days (indicated by the red bracket), and then to the infectious class for six days (illustrated with the red bracket). Infected individuals move to the recovered class and acquire immunity. Successfully vaccinated individuals move straight into the recovered class with full immunity.

From Figure 3.1 and the model description provided earlier, a system of difference equations representing the disease dynamics were derived as follows:

$$\begin{aligned}
S(t+1) &= S(t) - E_1(t+1) - \kappa, \\
E_1(t+1) &= \text{Binom}(S(t), p(t)), \\
E_2(t+1) &= E_3(t), \\
&\vdots \\
E_{10}(t+1) &= E_9(t), \\
I_1(t+1) &= E_{10}(t), \\
I_2(t+1) &= I_1(t), \\
&\vdots \\
I_6(t+1) &= I_5(t), \\
R(t+1) &= R(t) + I_6(t) + \kappa.
\end{aligned} \tag{3.2.2}$$

In Equation 3.2.2, κ is the number of individuals vaccinated during the campaign, represented mathematically as

$$\kappa = \begin{cases} 0, & t \neq t_c \\ \text{Binom}(S(t_c), p_v), & t = t_c \end{cases}$$

Also, t_c is the time of the vaccination campaign corresponding with the observed, p_v is the observed vaccinated proportion in the high school vaccination campaign, and $S(t_c)$ is the number of susceptible individuals at the time of vaccination.

3.2.2.1 Model assumptions

We assume 10 days of latency followed by 6 days of infectiousness, similar to Grais *et al.* (2007) and in accordance with the natural history of measles. Individuals become immune after the 16 days of infection and therefore move into the recovered class with lifetime immunity. Vaccinated individuals also acquire lifelong immunity.

Since the model describes school (closed population) measles epidemic dynamics, birth and death are not included in the model. Hence, there is no susceptible replenishment through birth or immigration, and no depopulation through emigration or death during the epidemic period. As a result, a constant population size N, is maintained throughout the epidemic.

The process through which an individual acquires the infection is assumed to be stochastic. Progression to the recovered class is, however, modelled as deterministic.

Since the appearance of a rash in a measles-infected individual is the visible indication of infectiousness, the index case is initialized into the $I(5)$ class, the reason being that the rash typically appears after four days, on average, and this person is known to be infectious during those four days. Individuals in the I class are, as a result, the driving force of the infection in the model. An individual therefore becomes exposed by coming in contact with an individual in any of the subclasses $I(1), I(2), \ldots, I(6)$.

3.2.2.2 Model parameters

The basic reproduction number $\mathcal{R}_0$, defined as the mean number of secondary infections produced by a single infectious individual in an entirely susceptible population, was assigned one of two values 12 or 18, based on frequently cited estimates in the literature (Anderson *et al.*, 1992; Guerra *et al.*, 2017).

From the chosen $\mathcal{R}_0$ values, we derived the transmission coefficient β - a quantity that encapsulates the number of contacts a susceptible person makes with an infectious person and the probability of getting infected through those contacts.

The transmission coefficient had a dimension equivalent to the duration of infectiousness that we assumed in the model and since the model had daily time-steps, we rescaled β to possess an equivalent daily rate. The rescaling was done as follows:

Let τ be the infectious period of measles and N, the constant population size. Then the expected number of infections per day is approximately $\frac{\mathcal{R}_0}{\tau}$. The daily probability of an infectious contact $p_c(t)$, at any time t, is therefore

$$p_c(t) = \frac{\mathcal{R}_0}{\tau N}. \tag{3.2.3}$$

Also, recall that the probability of infectious contact, at time t in the model, is formulated in Equation 3.2.1 as

$$p(t) = 1 - \exp\left(\frac{-\beta I(t)}{N}\right)$$

where $I(t)$ is the number of infectious individuals at time t.

At the beginning of the epidemic, we can equate Equation 3.2.3 and the results of $p(t)$ above to obtain

$$1 - \exp\left(\frac{-\beta I(t)}{N}\right) = \frac{\mathcal{R}_0}{\tau N},$$

and the transmission coefficient, in days, can therefore be approximated using

$$\beta = -\frac{N}{I(t)} \ln\left(1 - \frac{\mathcal{R}_0}{\tau N}\right). \tag{3.2.4}$$

Since, in the definition of $\mathcal{R}_0$, it is assumed that there is a single infectious individual in the population, it implies that $I(t) = 1$ and Equation 3.2.4 can be simplified to

$$\beta = -N \ln\left(1 - \frac{\mathcal{R}_0}{\tau N}\right). \tag{3.2.5}$$

The effective reproduction number, defined as the number of secondary cases produced by a single case in a partially immune population $\mathcal{R}_{eff}$, was assigned a value of 5, which was the number of secondary cases reported during the actual outbreak.

Knowing $\mathcal{R}_0$ and $\mathcal{R}_{\text{eff}}$, we estimated the initial susceptible proportion using the relation

$$s(0) = \frac{R_{\text{eff}}}{\mathcal{R}_0}.$$

We proceeded to use the results from the above to estimate the immune proportion of the population at the start of the epidemic as $r(0) = 1 - s(0)$. With $r(0)$ and $s(0)$, we then initialized the susceptible and immune classes of the model for simulation.

We introduced a scaling parameter Θ, which we used to account for the uncertainty surrounding the eventual number of epidemic cases. We assigned Θ one of two arbitrary values: 0.5 or 1. Θ was used to rescale the effective reproduction number, which represented the number of secondary cases produced by the index case in our model. We introduced Θ based on knowledge of the fact that during the course of an epidemic, transmission changes and hence, the model can either over- or under-estimate the number of expected cases. Herein, we will refer to Θ as the rescaling parameter.

As mentioned earlier, our objective was to estimate the cases averted, and to predict the cases remaining until the end of the epidemic, along with the corresponding time left. Therefore, we performed multiple scenario analyses to achieve our objectives. We employed the scenario analysis approach in order to: account for certain uncertainties surrounding the model, such as the value of $\mathcal{R}_0$, the expected number of cases, and so forth; and allow for counter-factual comparisons between scenarios involving vaccination and those without vaccination, so as to determine the cases averted - we calculated cases averted due to the vaccination intervention by finding the difference between vaccination and non-vaccination scenarios corresponding to the same parameter assumptions.

Scenarios were composed with a combination of: $\mathcal{R}_0 = 12$ or 18, with or without vaccination, and the rescaling parameter $\Theta = 0.5$ or 1. Table 3.1 compiles the scenarios described.

Table 3.1: Scenario definitions. Each column represents a different scenario labelled with code A to H. For example, scenario code A represents $\mathcal{R}_0 = 12$, vaccinate = No, and rescaling parameter $\Theta = 1$.

Code	A	B	C	D	E	F	G	H
$\mathcal{R}_0$	12	18	12	18	12	18	12	18
Θ	1	1	1	1	0.5	0.5	0.5	0.5
Vaccinate	No	No	Yes	Yes	No	No	Yes	Yes

For each scenario, 1000 simulations were ran. The results are presented in Section

3.3.

3.3 Results

3.3.1 Model evaluation

An assessment of the model's performance in predicting the epidemic duration and total cases is presented in this section.

3.3.1.1 Assessment of epidemic duration results

The following figures represent distributions of the epidemic duration simulation results per school - for the non-vaccination scenarios - compared with the actual observed epidemic duration of each of the epidemics described in Section 3.2.1.2.

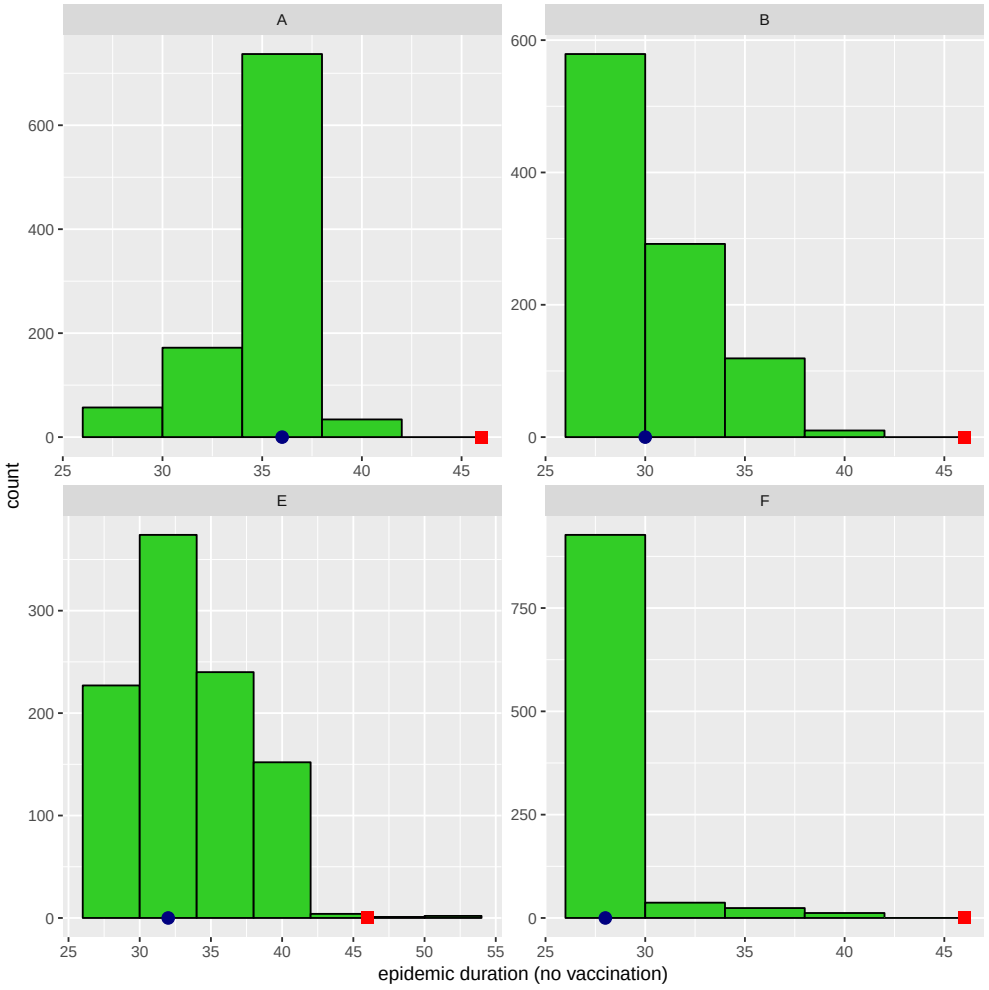


Figure 3.2: Epidemic duration simulation results for the epidemic in School A, Oxford (1929). Among all of the non-vaccination scenarios, the medians underestimate the observed epidemic duration. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. The scales are different on all four plots.

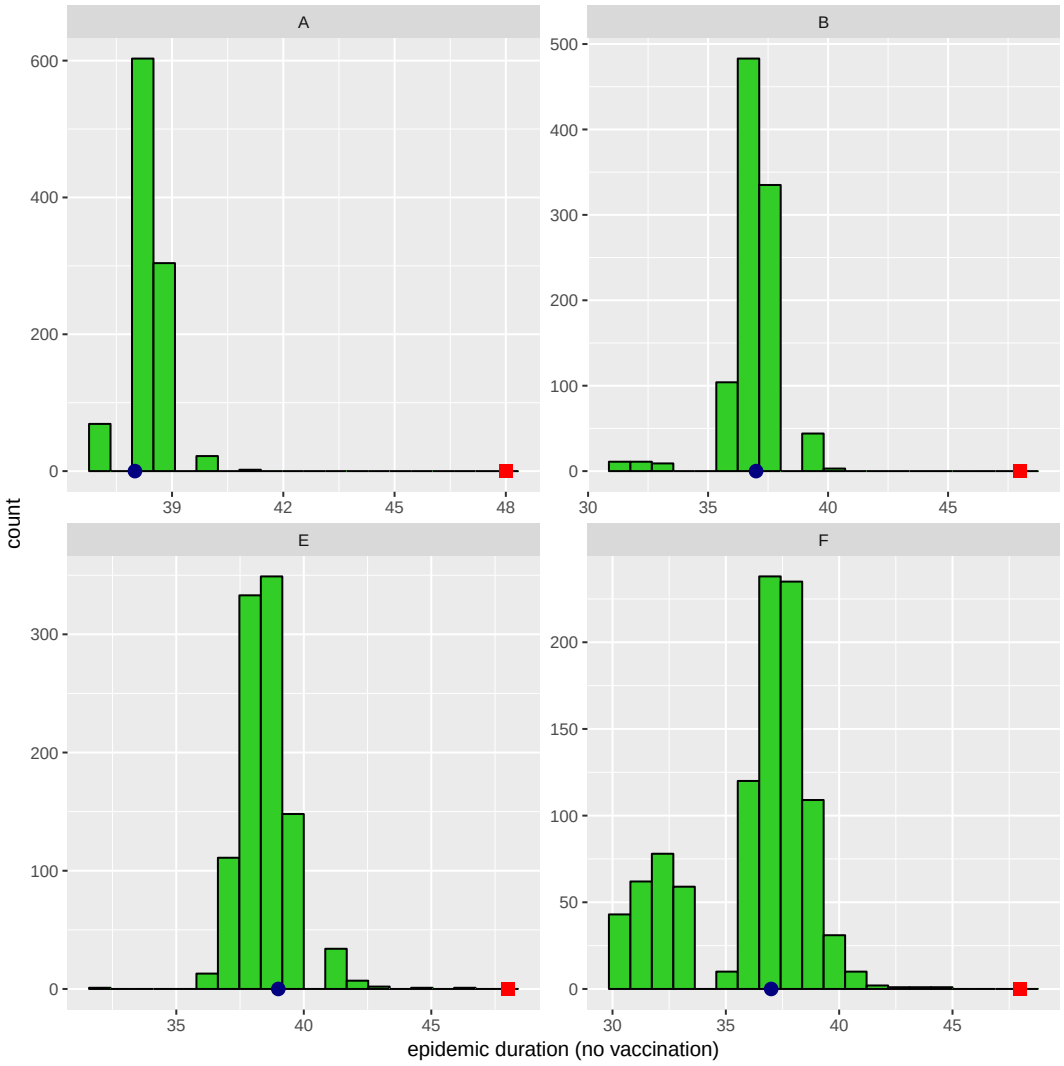


Figure 3.3: Epidemic duration simulation results for the epidemic in School B, Oxford (1930). The median of the results of all the non-vaccination scenarios underestimate the observed epidemic duration. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. The scales are different on all four plots.

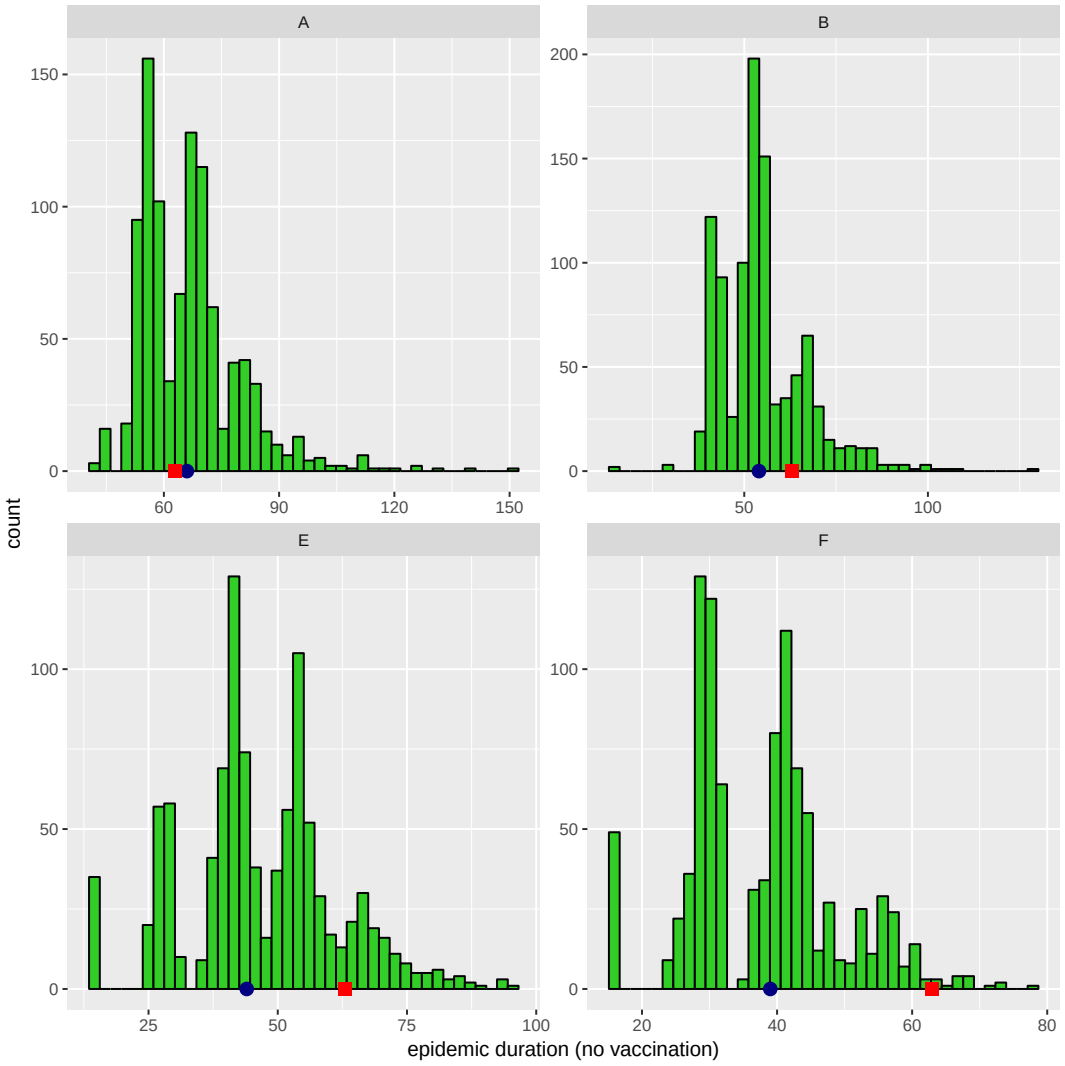


Figure 3.4: Epidemic duration simulation results for the epidemic in School B, Oxford (1934). The observed epidemic duration is underestimated by scenarios B, E, and F, but a close match is obtained with scenario A. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. The scales are different on all four plots.

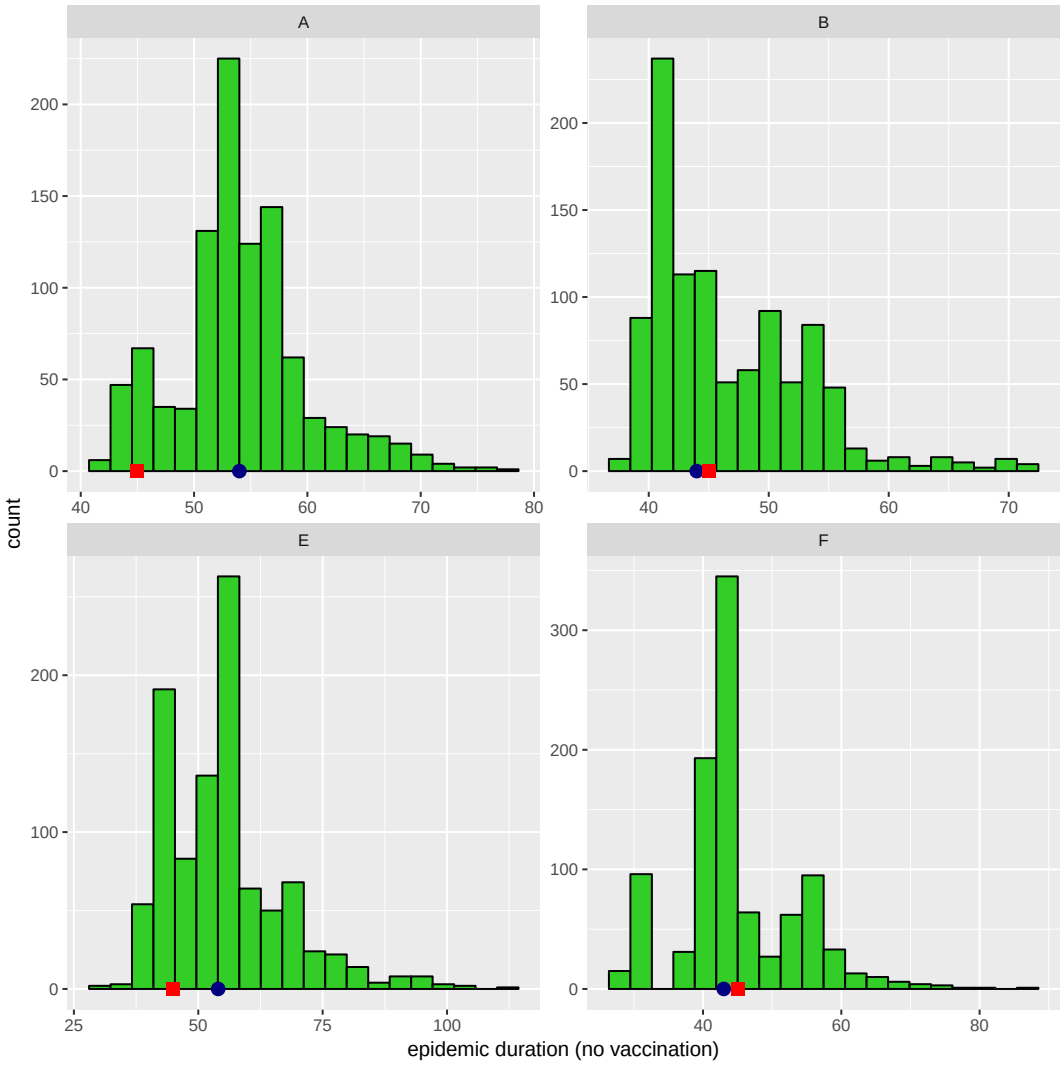


Figure 3.5: Epidemic duration simulation results for the epidemic in School A, Oxford (1938). Scenarios A and B overestimate the observed duration. An exact match is almost obtained by scenario E. Scenario F underestimates the observed. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. The scales are different on all four plots.

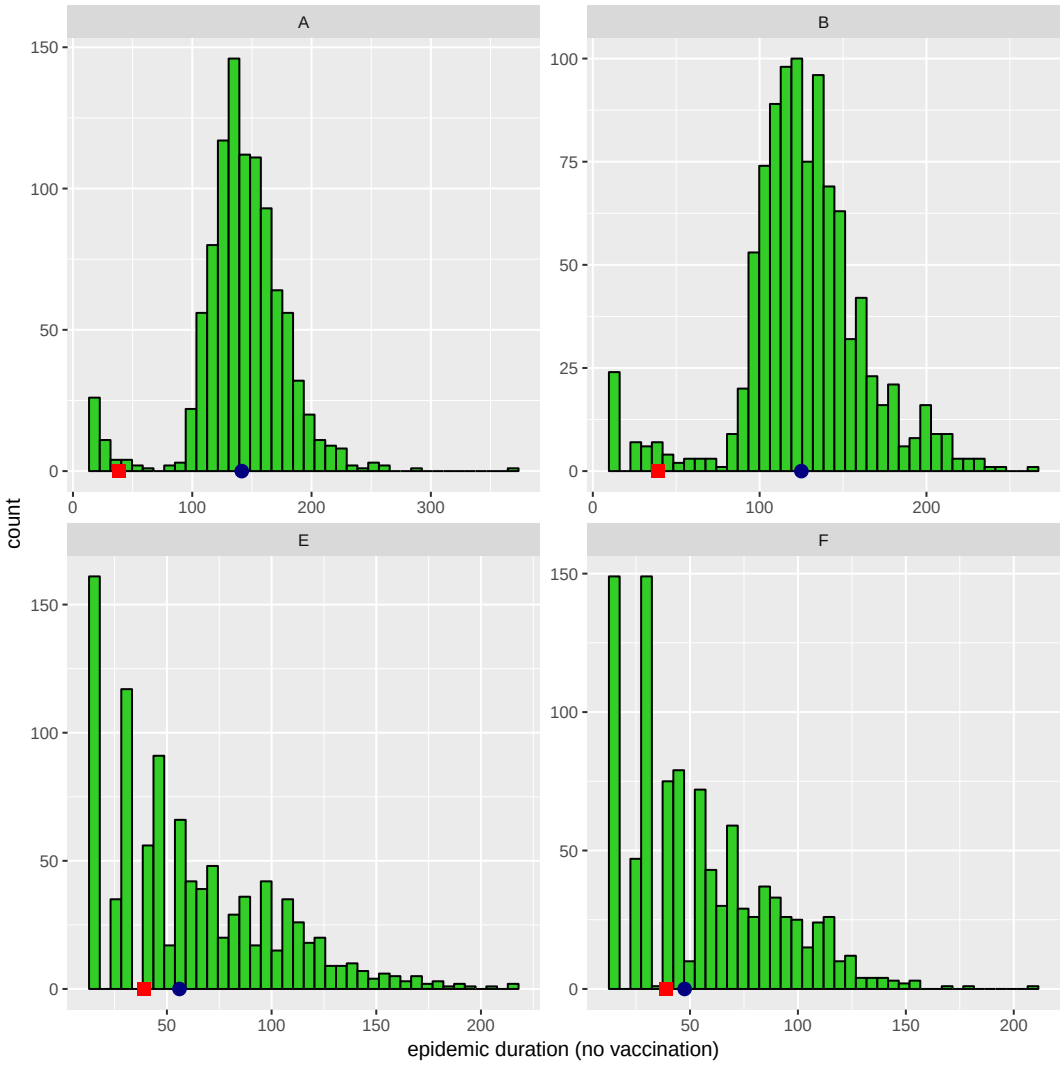


Figure 3.6: Epidemic duration simulation results for the epidemic in a boarding school, US (2003). All four scenarios overestimate the observed duration with scenarios E and F producing close matches. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. The scales are different on all four plots.

All the results presented are summarised in the table which follows.

Table 3.2: Summary of the model's performance in predicting the epidemic duration of various school epidemics. The proportion of simulation results that under-estimated the observed value, and the scenario that produced the closest median value are indicated.

Epidemic	Observed duration	Proportion under-estimated	Closest scenario	Median (closest scenario)
School A (Oxford, 1929)	47	99.9%	A	36
School B (Oxford, 1930)	49	100.0%	E	39
School B (Oxford, 1934)	64	76.2%	A	66
School A (Oxford, 1938)	45	35.3%	B	44
Boarding school (US, 2003)	39	18.8%	F	47.5

3.3.1.2 Assessment of simulated total cases

The model's performance in predicting the total number of cases at the end of each of the school epidemics, described in Section 3.2.1.2, is assessed in this section.

The figures that follow show the distribution of simulation results of epidemic total cases for the non-vaccination scenarios of each school epidemic, compared with the actual observed total cases.

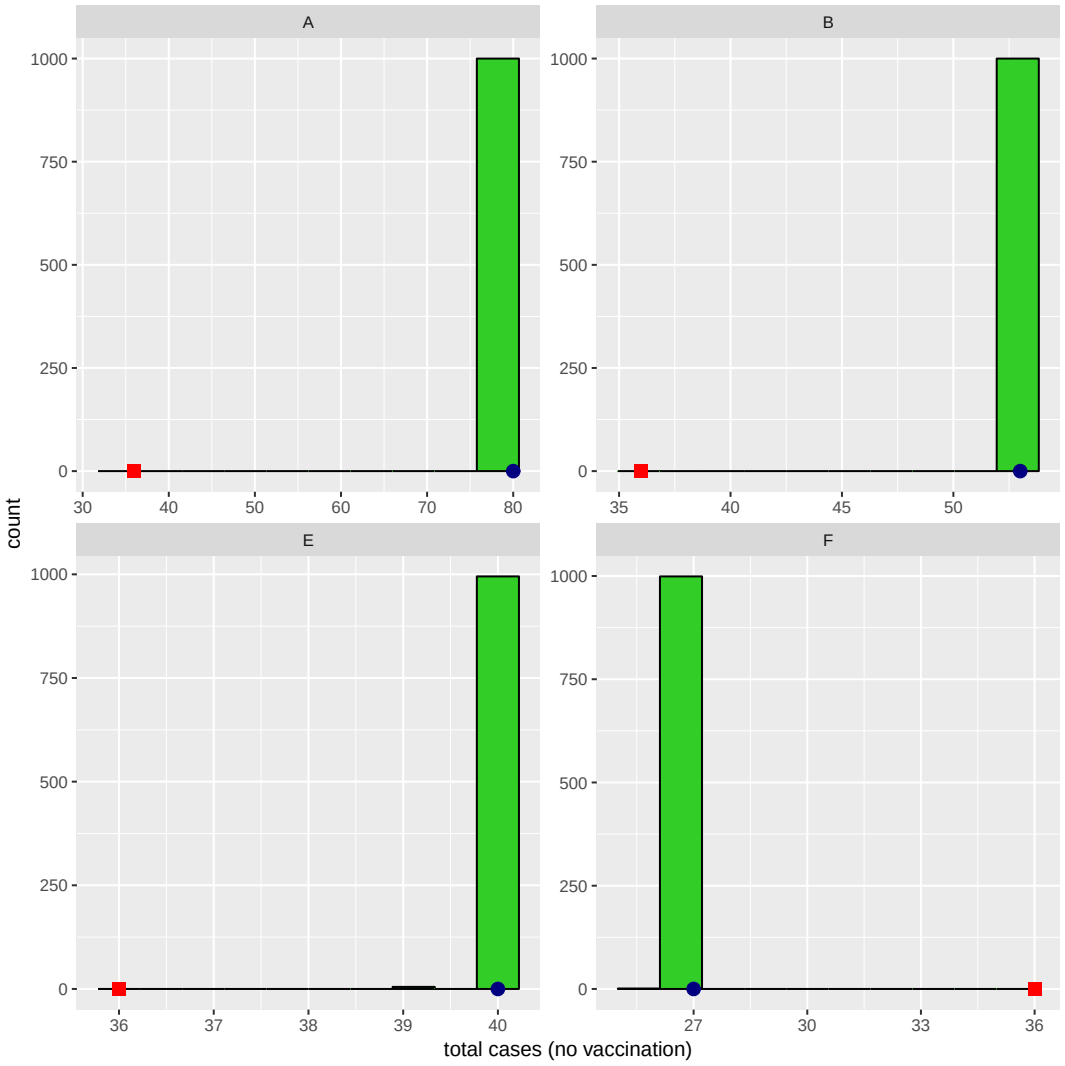


Figure 3.7: Simulation results of total cases for the epidemic in School A, Oxford (1929). In three of four scenarios, the model overestimates the total cases. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. All four plots have different scales.

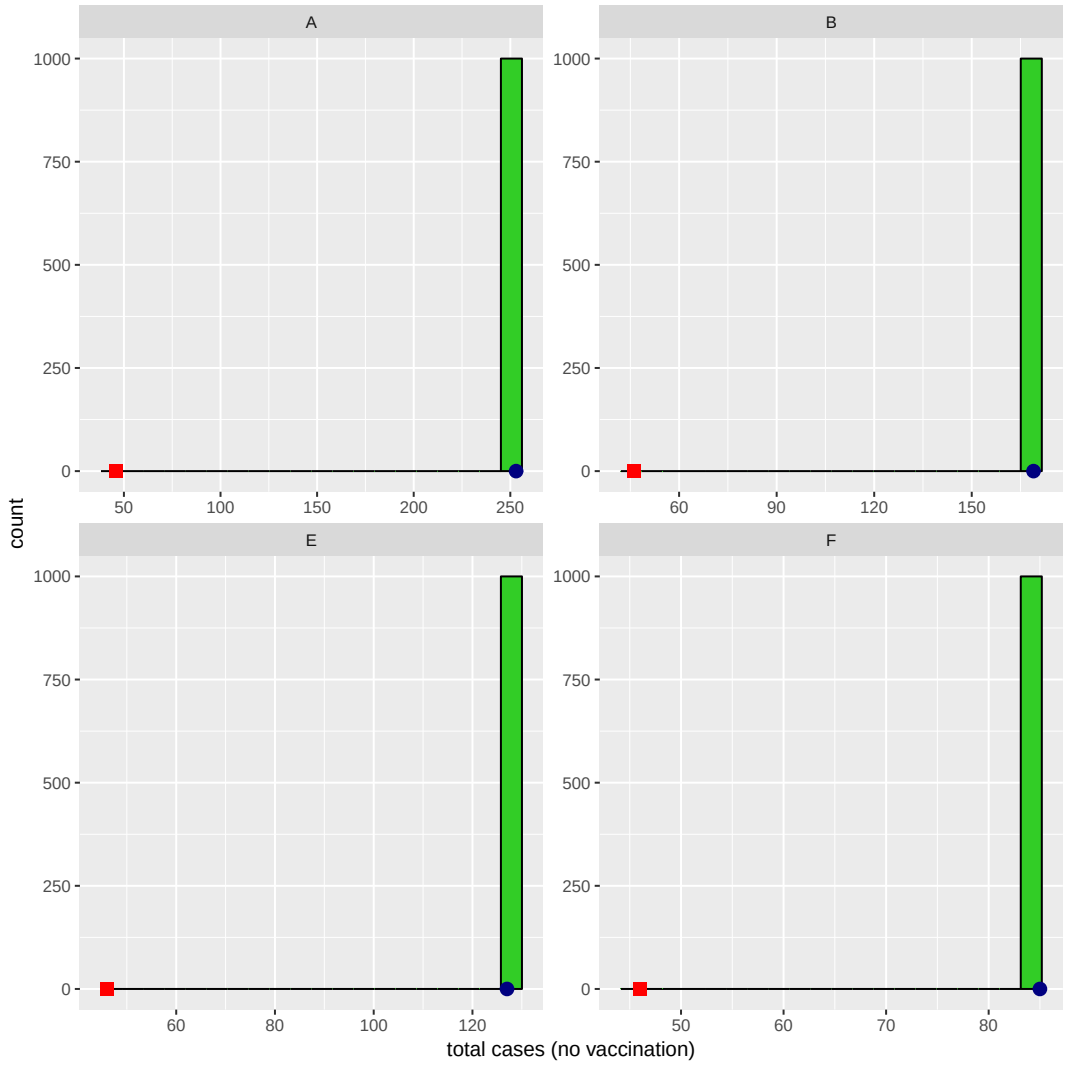


Figure 3.8: Simulation results of total cases for the epidemic in School B, Oxford (1930). The simulated median cases overestimates the observed in all four scenarios. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value.

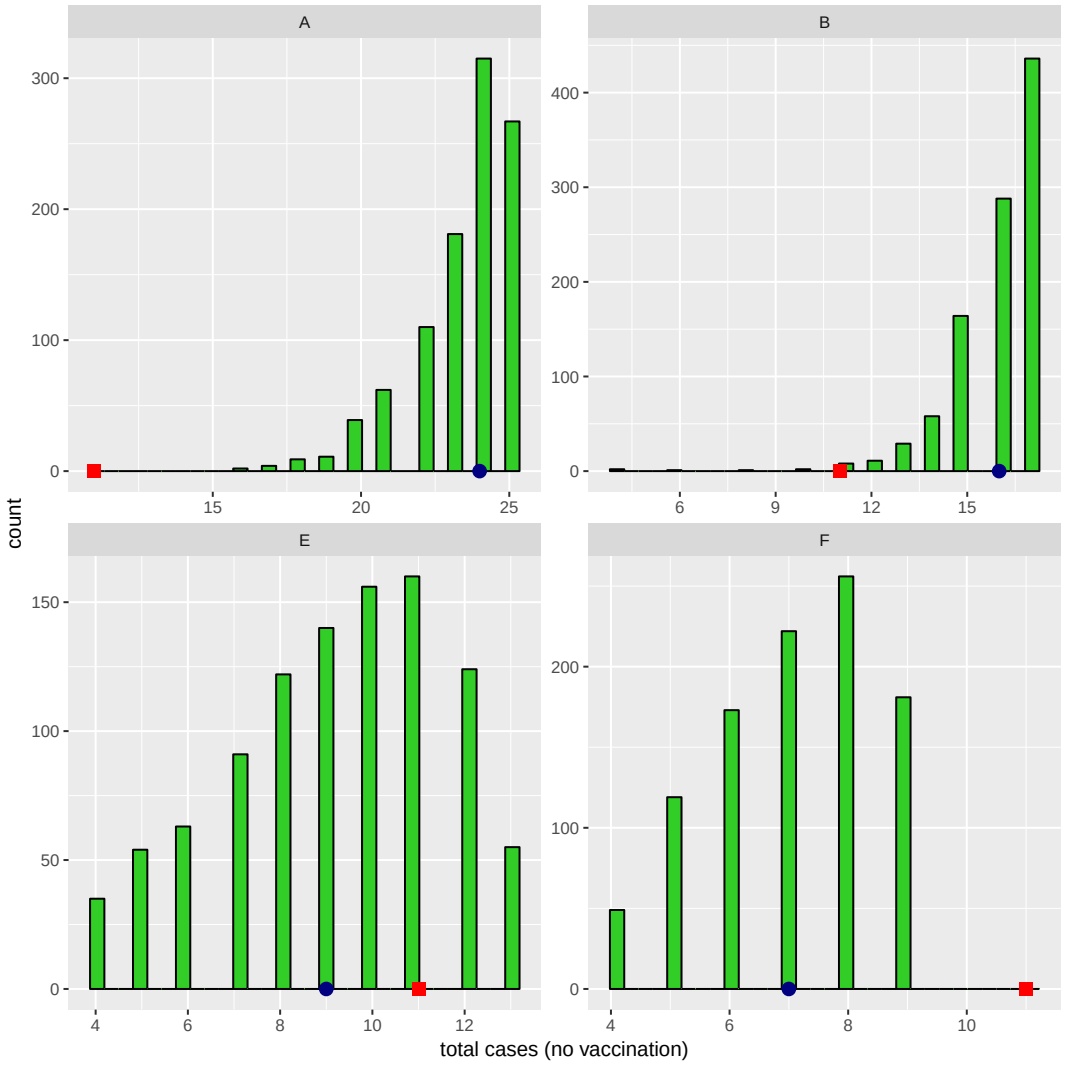


Figure 3.9: Simulation results of total cases for the epidemic in School B, Oxford (1934). Simulated median overestimates the actual total cases in two scenarios and underestimates in the remaining two scenarios. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. All four plots have different scales.

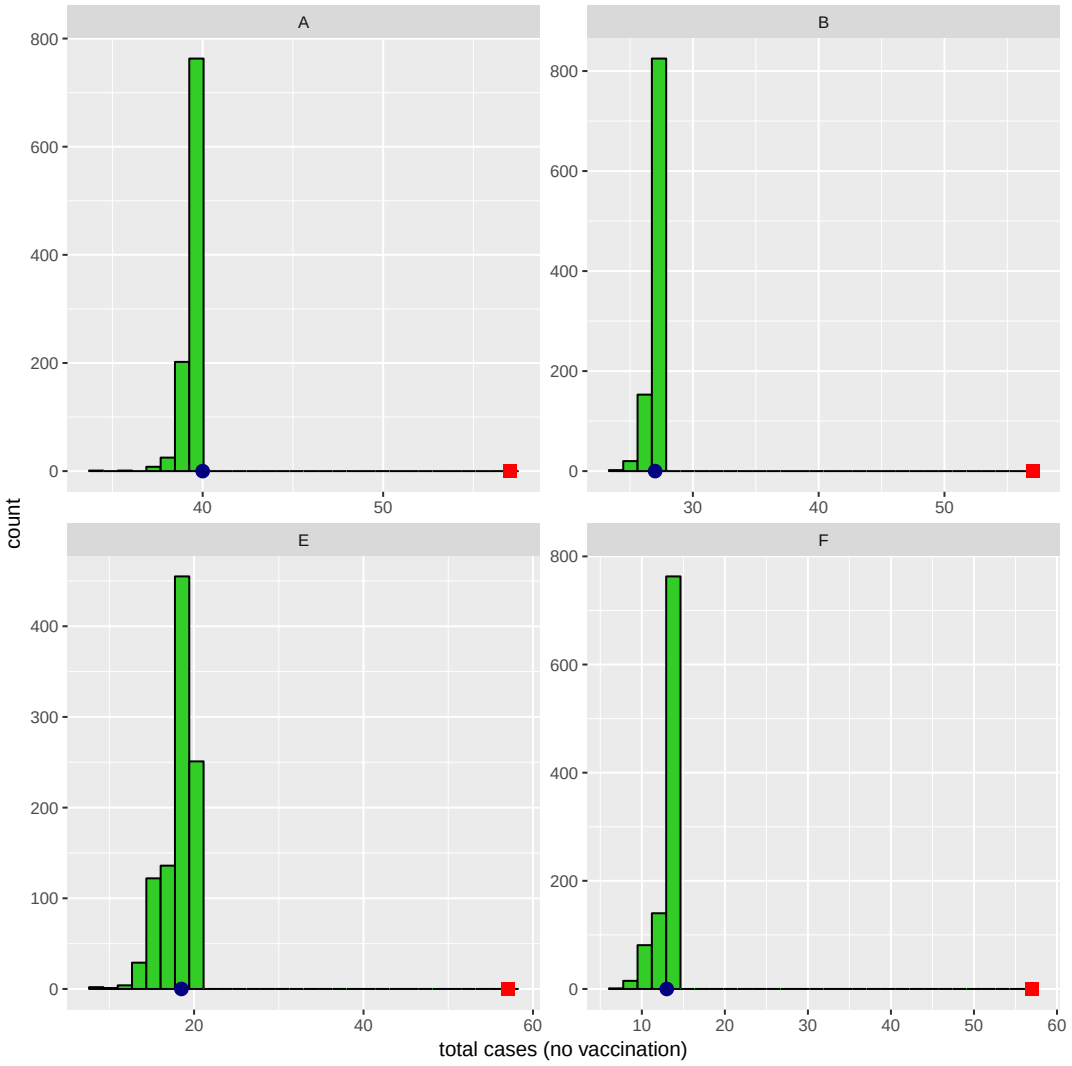


Figure 3.10: Simulation results of epidemic duration for the epidemic in School A, Oxford (1938)). The observed total cases is underestimated by all scenarios. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. All four plots have different scales.

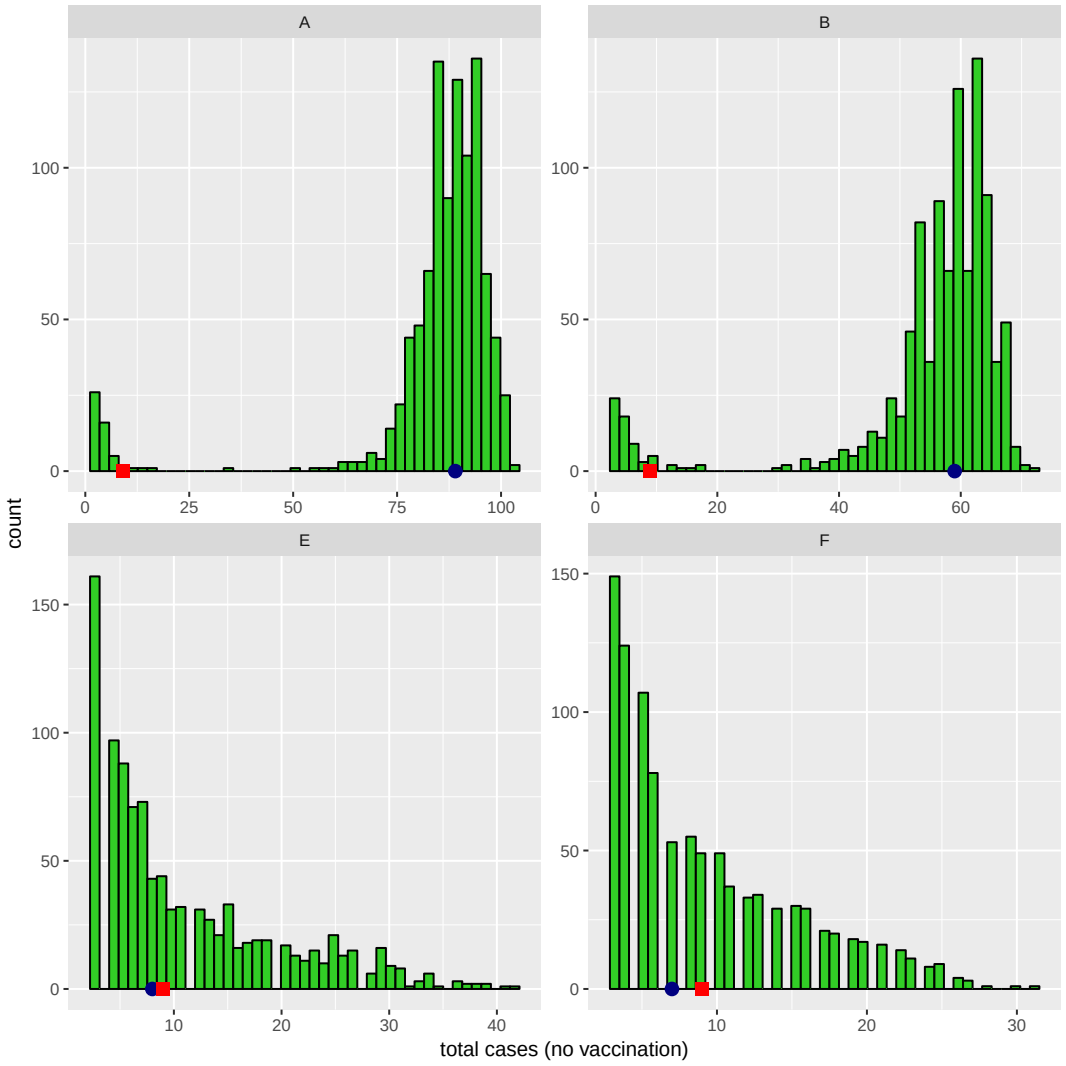


Figure 3.11: Simulation results of total cases for the epidemic in a boarding school, US (2003). Simulation median overestimates the observed cases. Scenarios E and F produce close matches. The median of the simulation results is indicated with the blue circle whereas the red square indicates the actual observed value. All four plots have different scales.

Table 3.3 summarises all the results of predicting the total cases presented here.

Table 3.3: Summary of the model's performance in predicting the epidemic total cases of various school epidemics. The proportion of simulation results that under-estimated the observed value, and the scenario that produced the closest median value are indicated.

Epidemic	Observed total cases	Proportion under-estimated	Closest scenario	Median (closest scenario)
School A (Oxford, 1929)	36	25%	E	40
School B (Oxford, 1930)	46	0%	F	85
School B (Oxford, 1934)	11	41.7%	E	9
School A (Oxford, 1938)	57	100%	A	40
Boarding school (US, 2003)	9	30%	E	8

3.3.2 Results of simulating the Stellenbosch school's outbreak

Figure 3.12 provides a box plot of the results of predicting the epidemic duration of the Stellenbosch all-boys high school outbreak, following 1000 simulation runs of each scenario.

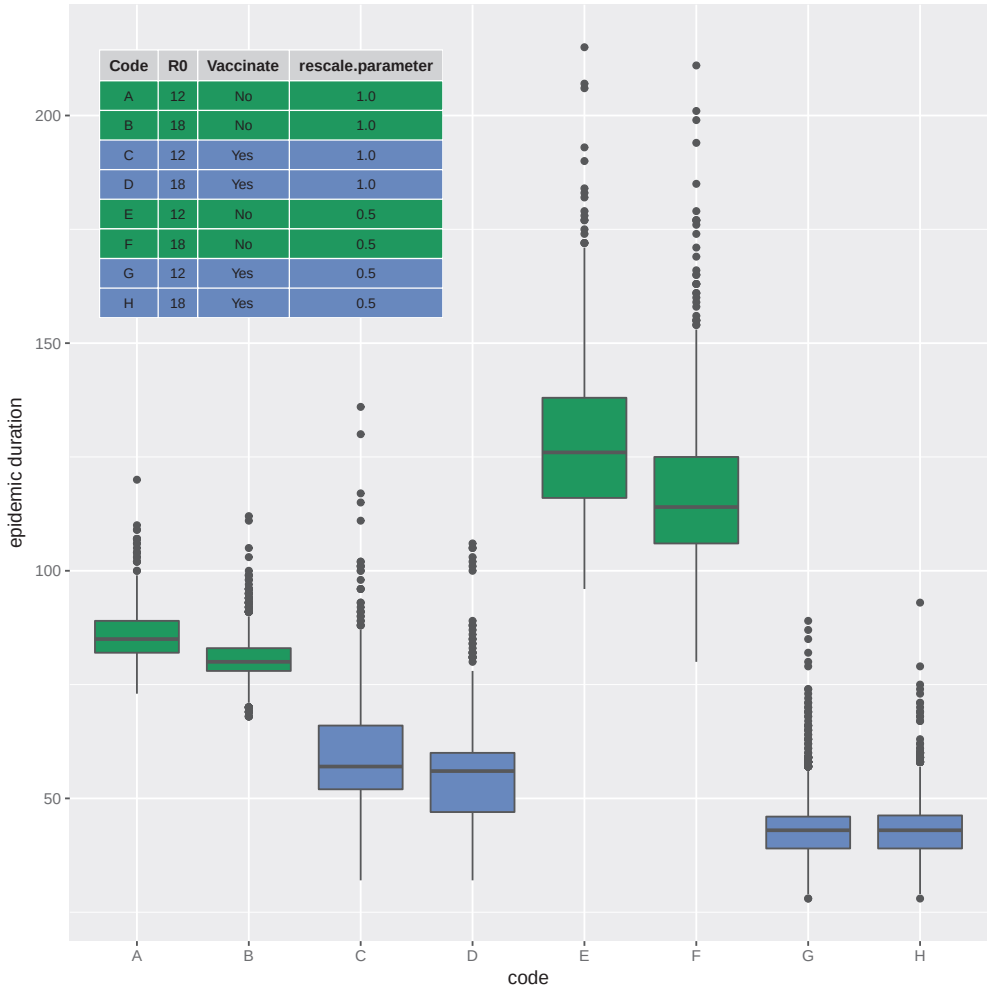


Figure 3.12: Summary of epidemic duration results from 1000 simulation runs per scenario. Scenarios with and without vaccination are color-coded with light-blue representing the former and light-green the latter. Scenarios without vaccination produce longer epidemic durations compared to those with vaccination.

It is immediately obvious from Figure 3.12 that with vaccination, and $\Theta = 0.5$, an $\mathcal{R}_0 = 12$ or 18 generates the two least epidemic durations (Scenarios G and H respectively). In the absence of vaccination, even if the second generation is rescaled by 0.5, an $\mathcal{R}_0$ of 12 or 18 produces the two longest durations of outbreak (Scenarios E and F respectively).

Similarly, a summary of the results of the simulated total cases of the Stellenbosch all-boys high school outbreak is illustrated with a box plot as shown in Figure 3.13.

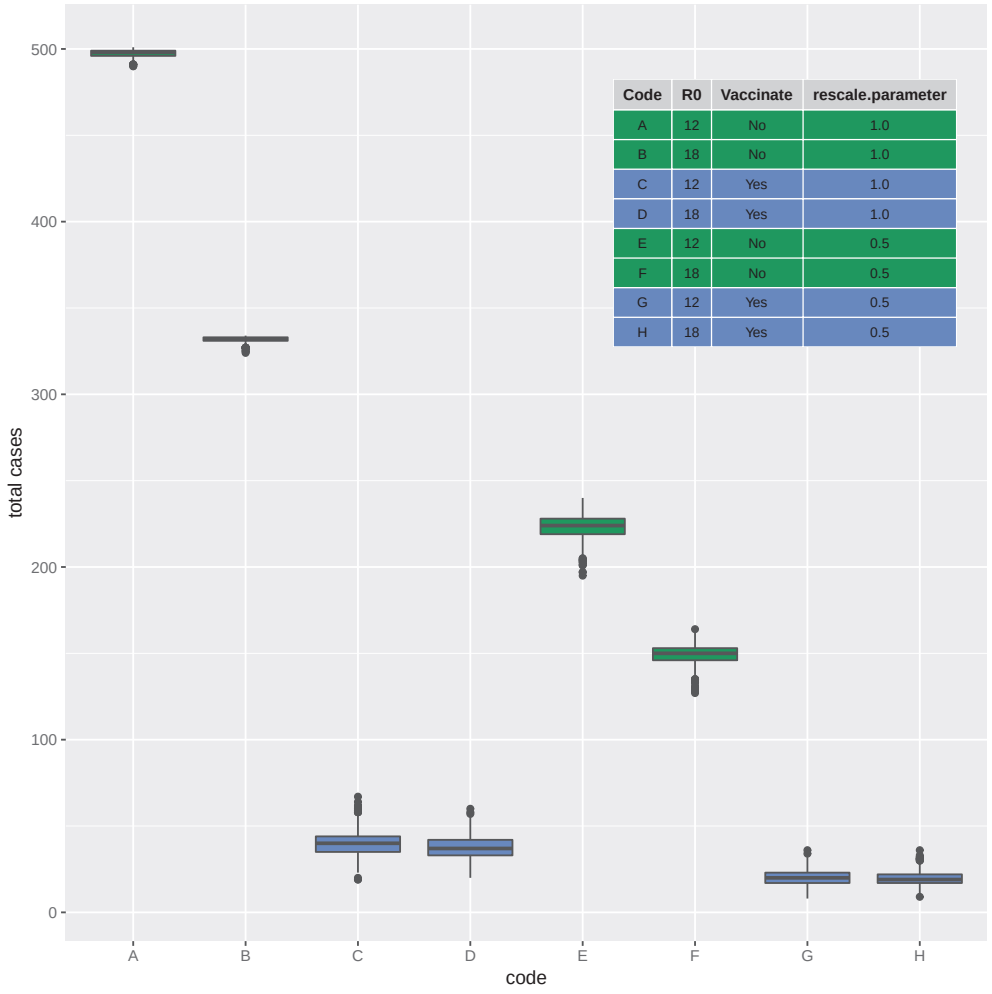


Figure 3.13: Summary of results of total cases of each scenario per 1000 simulation runs each. The light-green plots represent scenarios without vaccination, whiles those with vaccination are represented with light-blue. Scenarios without vaccination produce higher numbers of cases compared to those with vaccination.

From the results, scenario A yields the highest number of cases, while the least number of cases results from scenario H. The total number of cases from C and D are nearly undifferentiated, and in a similar fashion, those from G and H. The four vaccination scenarios (C, D, G, H) produce the least total cases.

We aggregated the results of the epidemic duration and total cases across all scenarios and evaluated the mean and range of the combined results. This is presented in Table 3.4.

Table 3.4: Median and mean total cases and epidemic duration (in days) after aggregating the scenarios. Ranges are reported in round brackets.

	Median		Mean	
Quantity	Vaccination	No vaccination	Vaccination	No vaccination
Total Cases	27 (8 – 67)	282 (127 – 501)	29.254	300.4572
Epidemic duration	47 (28 – 136)	96 (68 – 215)	50.592	102.7935

It is immediately obvious that with vaccination, both fewer cases and shorter epidemic durations are realized, compared to the scenarios without vaccination.

Next, we present the results of estimating the number of cases averted, by finding the difference between vaccination scenarios and their counterfactuals.

3.3.2.1 Cases averted

Table 3.5 shows estimates of cases averted following the vaccination campaign. According to the model with $\mathcal{R}_0 = 18$ and $\Theta = 0.5$, the least median number of cases are averted. This represents the difference between scenarios F and H. Further, the difference between scenarios A and C, which assume an $\mathcal{R}_0$ of 12 and a rescaling parameter $\Theta = 1$, yields the highest median cases averted.

Table 3.5: Cases averted per 1000 simulations of each scenario, computed as the difference between comparable vaccination and non-vaccination scenarios. "Scenario difference" is to be read as non-vaccination scenario minus vaccination scenario. The least and highest median number of cases averted are (F−H) and (A−C) respectively. Ranges are reported in the round brackets. The overall cases averted, highlighted in grey, represents the difference between the aggregated results from the vaccination and non-vaccination scenarios.

Scenario difference			Cases averted	
Code	$\mathcal{R}_0$	Rescale parameter	Mean	Median
A−C	12	1	457.596	458 (423 − 482)
B−D	18	1	294.094	295 (264 − 314)
E−G	12	0.5	203.347	204 (159 − 232)
F−H	18	0.5	129.776	131 (91 − 155)
Overall	-	-	271.2032	255 (60 − 493)

In the following section, we present the results of our predictions of the expected cases and time remaining, following the campaign, until the epidemic ended.

3.3.2.2 Time, and cases remaining till epidemic die-out.

The time difference (in weeks) between the date of rash onset of the index case, 16th January 2017, and the date of the last simulation ran, 23rd February 2017, was found to be 28 days. With that, predictions of the time remaining for the epidemic and the expected remaining cases after the vaccination campaign, which achieved a coverage of $> 93\%$, were calculated. Total cases at the time of the last simulation run was 23.

The aggregated vaccination scenarios yielded a median of 15 (range $1 - 33$) remaining cases and 4 (range $1 - 16$) remaining weeks. Table 3.6 shows the results per vaccination scenario.

Table 3.6: Median time to end of outbreak (in weeks) and remaining cases per vaccination scenario following vaccination. Codes C, D, G, and H represent the vaccination scenarios. Ranges are reported in the round brackets. The least and highest number of cases and weeks remaining are scenarios H and C, respectively. The overall summary is highlighted in grey, and represents the median and range across the aggregated results from the vaccination scenarios.

Scenario code	$\mathcal{R}_0$	Rescale parameter	Cases remaining	Weeks remaining
C	12	1	15 $(4-33)$	5 $(2-13)$
D	18	1	15 $(4-28)$	4 $(2-12)$
G	12	0.5	5 $(1-9)$	3 $(1-7)$
H	18	0.5	5 $(2-7)$	1 $(1-4)$
Overall	-	-	15 $(1-33)$	4 $(1-16)$

3.4 Discussion

The main objective of this chapter was to evaluate the impact of a reactive vaccination campaign organised by the Western Cape Department of Health in a school in South Africa. According to the results obtained, the campaign prevented a considerable number of cases from occurring. It is known that vaccination effectively reduces measles transmission and spread (WHO, 2017; Yeung *et al.*, 2005) and hence, this outcome was anticipated. Information on how many cases were averted is useful in quantitatively appraising reactive vaccination campaigns. Our study estimates that more than two hundred and fifty cases were prevented. The results of this chapter extends the growing literature on the impact that outbreak response vaccination has on measles epidemics. These results are, however, to be interpreted within the confines of some strong assumptions made in the model. These assumptions are discussed as potential limitations of the study.

The secondary goal of this chapter was to provide estimates of how much time would pass between the vaccination campaign and the end of the epidemic as well as how many cases to expect in the said time. The observed epidemic duration was consistent with the model's overall minimum expected epidemic duration. According to the model, the epidemic was expected to end within $1-16$ weeks. In fact, the epi-

demic had actually ended by the time we were running our analyses, a few weeks after the campaign. The model also predicted a range of $1 - 33$ remaining cases. The model only focused on the number of cases among learners in the school, and we do not know how many of the cases post-vaccination actually occurred in this population, but we are certain the observed remaining cases was captured within the predicted bounds of the model.

The results of the secondary goal would have been important for planning ahead. Even though the results were useful for communication, they were not timely enough to be useful as a real-time tool for assessment and decision-making. Nevertheless, the methods we employed to achieve the results can be improved for studying closed-population measles epidemics in real-time. For modelling outcomes in fast outbreak settings like that of measles it is preferable that the model be ready sometime in advance to facilitate real-time predictions and communications.

Measles outbreak response strategies dictate that the highest possible vaccination coverage attainable is prioritized over every other outcome, but studies have shown that in some cases, speed of response is more important (Ferrari *et al.*, 2014; Grais *et al.*, 2007). In this light, the reactive campaign organized in the high school was fast relative to the epidemic progression and achieved a high coverage ($\approx$ 93%). The estimate of cases averted obtained from this study is, therefore, testament to the importance of response speed and high vaccination coverage. This study did not, however, explicitly study how the timing of the intervention influenced the results.

The mere occurrence of the outbreak in the school after a several years of no outbreaks deserves further investigation. It is clearly an indication of the possible existence of a pool of susceptible learners, built up over time. Even though there were only 5 confirmed cases at the time of the campaign, the 23 cases that followed marked its climax and points to the existence of high susceptibility in the population. A susceptible pool builds up due to non-uptake of the vaccine, often for religious reasons; immunization failure among some of the learners; loss of immunity; and lack of previous exposure to the infection (WHO, 2017). This outbreak warrants the need to investigate immunity levels among learners, not only in the affected school but in the whole province. Vaccination uptake should be encouraged to ensure sero-conversion among the learners who may have failed previous vaccination attempts. This is so that the herd immunity of measles is achieved to protect against future outbreaks. If measles elimination is a goal in the province, this is an important next step.

3.4.1 Potential limitations

We assumed the range of widely cited $\mathcal{R}_0$ values of 12 and 18 in formulating the multiple scenarios. It is known that $\mathcal{R}_0$ varies widely and is context-specific (Guerra *et al.*, 2017). Future extensions of this study will include a broader range of $\mathcal{R}_0$ values to compensate for the uncertainties surrounding the actual value.

Another potential limitation of our modelling approach was the assumption that the vaccine had 100% efficacy and that the immune response was instantaneous, thereby neglecting the time delay until an immune response kicks in. The former assumption is contrary to other modelling studies in developing countries which assume lower efficacy rates (Grais *et al.*, 2007; Lessler *et al.*, 2016). However, considering South Africa is a hybrid of developing and developed provinces, with the Western Cape province being developed and highly-resourced in the health sector, we made vaccine efficacy assumptions relevant to high-resource settings. This assumption followed after we had carefully examined the reasons behind loss of efficacy in measles vaccines, and compared them with the characteristics of the population we were modelling, and the current state of affairs at the Western Cape Department of Health. This assumption is realistic as some measles vaccine efficacy studies in developed countries have estimated this quantity to be 100% or close to it (Uzicanin & Zimmerman, 2011). Vaccines remain efficacious when they are stored in subzero temperatures until they are administered. Also, the young age of vaccinees, associated with the possession of passively acquired maternal antibodies mentioned in Chapter 1 is another hindrance to efficacy (Aaby *et al.*, 1990). These criteria were not applicable to our study since the individuals in consideration were predominantly teenagers. Also, since the Department of Health in the Western Cape province is well-funded, it was our reasoning that the cold chain is well-managed to ensure the vaccines are stable and efficacious. We, therefore, consider it unlikely that the assumption about vaccine efficacy substantially biased our estimates. The assumption regarding the timing of the immune response, however, is a strong assumption with the potential to influence the results. Hence, the results obtained here are to be interpreted with caution. Future studies will investigate the effect of this delay.

The model did not fare too well in predicting the dynamics of other school epidemics during our validation. This may be due to a mismatch between the actual reproduction numbers ($\mathcal{R}_{eff}$ and $\mathcal{R}_0$) in those epidemics, compared with what was assumed in the model scenarios, or differences in other parameters. We intend to revise our assumptions and perform further assessments to ensure our predictions are more realistic for use in future outbreak situations. As a consequence of this limitation, we report our estimates with some level of caution. The approach we employed can be refined and

made more rigorous by relaxing some of our assumptions to improve the model's performance.

Additionally, a major constraint when modelling outcomes during outbreaks is time and uncertainty in decision-making and this was no different in our case. As a result, some of the decisions taken during the modelling process may have been made hastily. A choice needed to be made between making speedy and simplistic assumptions to account for uncertainties, or being more thoughtful but taking a longer time to respond to the outbreak. We chose the former. We, however, tried to keep the uncertainties under check by making assumptions specific to this particular outbreak.

3.4.2 Information brokerage

A secondary goal of outbreak response, according to the WHO guidelines is community awareness as prevention (WHO, 2009). In order to help achieve this goal at the time, several strategies were utilized. We will now proceed to discuss them.

3.4.2.1 Measles information sheets

Two kinds of information sheets were designed and distributed. The first, found in Appendix A.1, was distributed to the general public both electronically and on notice boards across the Stellenbosch university campus, as well as the surrounding high schools. The sheet contained information about the natural history of the disease as well as its means of spread. It also explained the advantages of being vaccinated against measles and gave instructions on what should be done in case an individual was found showing suspicious symptoms. We included information on the various clinics and pharmacies offering vaccination and case management services in Stellenbosch and its environs. Each of these clinics and pharmacies were contacted to ascertain the services they provided and associated costs. Pertaining to the costs, even though the government facilities provided free services, the private ones offered them at variable costs. All this information was translated into two of the most widely spoken languages in the Western Cape province: Afrikaans (translated by a SACEMA associate) and isiXhosa (translated by the Stellenbosch University Language Center).

The other sheet, included in Appendix A.2, was specific to the affected high school. It contained information on how many cases had been prevented, and how the epidemic was expected to progress in terms of the cases and time to expect until the epidemic had subsided following the vaccination campaign.

All school principals in the vicinity of the affected school were emailed with copies of the sheets in Appendix A.1 for further proliferation. The latter sheet, A.2, was sent to the principal of the affected high school under study and forwarded to the school community. The school recognized our efforts and this was acknowledged on their website. The following link directs you to the acknowledgment: https://www.paulroos.co.za/masels-measles/.

3.4.2.2 Web application for learners

We developed a web application using the R statistical language's (R Core Team, 2017) RShiny package version 1.0.5 (Chang *et al.*, 2017). The application was deployed to the school and neighbouring schools, after it was hosted on a landing web page. The application was designed to give the school's staff and learners insights into how the field of epidemiological modelling is employed to control the spread of infections. It was our hope that the application would boost mathematics pedagogy in the school as well as the learners' interests in the subject of mathematics. We also hoped to introduce the learners to epidemiological modelling as one of the many specializations in mathematical science.

The application consists of various tabs for learning about this specific measles outbreak including a tab under which the user can tweak certain parameters which determine the size and duration of the outbreak including: number of learners in the school (N), number of learners infected by the index case R_{eff}, timing of the outbreak response vaccination after diagnosis of the first case with a rash, and vaccination coverage achieved through the campaign. Another tab shows what could have happened without the vaccination in the best and worst case scenarios. The final tab explains a variety of concepts including: the plots displayed in the other tabs as well as the terminology used, and how the calculations were done in non-technical terms. This tab also describes some other applications of epidemiological modelling and how some other SACEMA (South African Center of Excellence for Epidemiological Modelling and Analysis)-affiliated researchers have successfully used mathematical modelling tools in disease management and control across the world.

The following is a link for accessing the application: http://measles.sacema.org/. Additionally, the implementation code for this chapter is hosted on a private github repository, available upon the reader's request.

Chapter 4

Conclusion

This thesis has answered two main questions, with an overarching aim of contributing to what is known about the two measles parameter estimation methods, and outbreak response vaccination impact assessment.

We started off in Chapter 1 by engaging with the ongoing scholarly conversation about measles dynamical modelling, methods for parametrizing measles models, and the impact of outbreak response vaccination. In Chapter 2, "Evaluating two methods for estimating measles parameters", we demonstrated that the performance of two well-known methods of fitting measles data, namely, the time series susceptible-infected-recovered model (Bjørnstad *et al.*, 2002) and the removal method (Ferrari *et al.*, 2005), depends on the assumptions underlying the data generation process, degree of spatial aggregation of the cases, the artificial methods employed to put the data in a form usable for each of the two models, and the parameter being fitted. In Chapter 3, "Measles outbreak in a high school in South Africa: assessing the impact of a reactive vaccination campaign", we assessed the impact of a reactive campaign organized by the Western Cape province's Department of Health in response to an outbreak in an all-boys high school in the town of Stellenbosch in South Africa. We found the reactive campaign to have been successful in preventing hundreds of cases and reducing the duration of the epidemic.

In-depth discussions of our findings have already been provided in each of the earlier chapters. Here, we suggest some implications of the findings of this thesis on two main aspects of the existing literature, namely, measles model parameterization, and outbreak response vaccination.

4.1 Implications of findings on measles model parameterization

Measles data fitting techniques, like the two we evaluated in Chapter 2, are usually formulated for well-curated, high quality data sets. Such data sets are not readily available in developing countries. When the data being fitted does not conform to that required by these types of models, decisions have to be made in the data modification process. Biases resulting from these modifications and the fitting process should be documented to guide how the estimates are interpreted.

The results of Chapter 2 imply that when parameters are fitted from large degrees of spatially aggregated data, the estimates are biased and hence the model's performance diminishes. Therefore, the results of estimates at this level of aggregation should be interpreted with caution.

The removal method tends to estimate S_0 with less bias, even when each 0 in the case time series is replaced with 1, but exaggerates $\mathcal{R}_{eff}$ in non-naive populations. The $\mathcal{R}_0$ estimates from the naive populations are biased but at a lower order of magnitude compared with those from the non-naive populations. The biases in the $\mathcal{R}_{eff}$ estimates are, therefore, driven by the fact that the population is non-naive and, therefore, violates the removal method's assumption of naiveness. This indicates that it is not advisable to fit non-naive populations with the removal method. When applying the removal method to data interspersed with zero cases, care should be taken in interpreting $\mathcal{R}_{eff}$ and $\mathcal{R}_0$, especially after the zero cases have been replaced with alternatives.

In future assessments, other approaches of dealing with zero cases should be tested, and the biases compared, in order to determine the approaches that produce unbiased estimates.

4.2 Implications of findings on outbreak response vaccination

The work done in Chapter 3 provides, at the least, a simple model for assessing outbreak response vaccination in closed populations. The limitations of the model have already been addressed in Section 3.4, including, most importantly, the assumption about vaccine efficacy, as it has to be as context-specific as possible in our case.

As epidemic data becomes available, the model should be parameterized and readied for future outbreaks as elimination is being targeted. We mentioned in our discussions in Section 3.4 that time was a constraint in our modelling response to the outbreak. It is therefore important that models should be developed in advance, in anticipation of

future outbreaks.

The outbreak response strategy mounted by the Department of Health was a success, since it averted a median of over 250 cases, in the confines of our modelling assumptions. Even though we did not explicitly investigate how the speed of the response affected the outcome, it is likely that the response time, coupled with the vaccination coverage achieved, were the reasons for the short epidemic. There are valuable lessons in this strategy for international efforts at achieving measles elimination, that is, to improve response time during outbreaks in closed populations - aggregation of school-going children drives measles epidemics, and if response is delayed, school epidemics can serve as hot beds for national outbreaks, thereby marring efforts at achieving local measles elimination.

Finally, it is known that in countries where vaccination has been well implemented, there is a shift in the mean age of infection from below 5 years to the teens. The outbreak in Stellenbosch provides clear evidence of the need to boost surveillance and immunity of school-going children to reduce the rate of build-up of susceptible pools. Health budgets should include measles serological surveys to help bridge the immunity gaps to facilitate achieving the measles elimination goal.

Appendices

Appendix A

Here, we present the information sheets described in the summary section of Chapter 3.

A.1 Information Sheets in English, Afrikaans, and isiXhosa

Measles outbreak in Stellenbosch

For more information, email **measlesinfo@sacema.org** or call 021 808 2589.

A SACEMA Information Sheet (Issued 08 Feb 2017)

Measles has been confirmed in Stellenbosch, and the number of cases is increasing. **What do you need to know?**

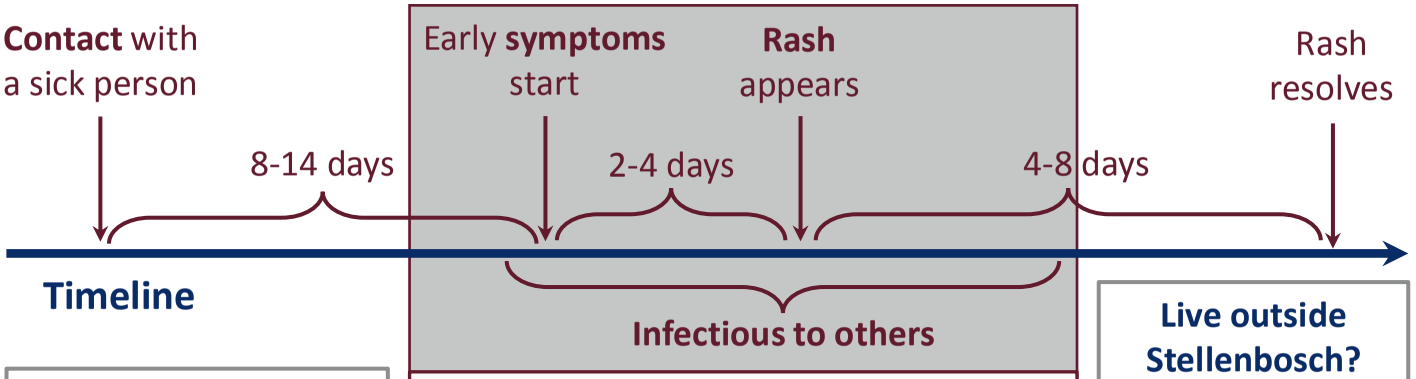
Vaccination reduces your risk of getting measles. If vaccinated, you may still get sick, but your symptoms will be less severe.

What can I do?

If you or someone you know has symptoms of measles, call your doctor or local clinic. Notify the health care staff that you may have measles before your visit. Stay away from others until four days after the rash appears. Keep sick children home from school until four days after the rash appears.

If you do not have symptoms, vaccination is recommended – especially if you have never received the vaccine, if you have been around someone with measles, or if you have a weakened immune system. Pregnant women should not get vaccinated.

Local government clinics are offering infant vaccinations and booster shots for free. Local pharmacies are offering measles vaccination for R100-R220.



How does it spread?

Measles is passed through **contact**, which can include touching someone, breathing the same air, or touching contaminated surfaces. Symptoms appear 8-14 days after contact with a sick person.

When can someone with measles infect others? One to two day before early symptoms start and up to four days after the rash appears.

Live outside Stellenbosch?

You could still be affected. Measles is one of the most contagious diseases and spreads quickly!

What are the symptoms?

Early **symptoms** are similar to a head cold and may include coughing, sneezing, and red eyes, accompanied by a high fever. After 2-4 days, a rash develops. The **rash** spreads from the head and neck to the whole body.

SACEMA DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis

NRF National Research Foundation

UNIVERSITEIT STELLENBOSCH UNIVERSITY

science & technology Department: Science and Technology REPUBLIC OF SOUTH AFRICA

Measles outbreak in Stellenbosch

For more information, email
measlesinfo@sacema.org or call 021 808 2589.

A SACEMA Information Sheet (Issued 08 Feb 2017)

Local government clinics are offering infant vaccinations and booster shots for free.
Find a clinic near you.

Cloetesville Community Health Clinic
Tennantville Street, Stellenbosch
021 883 2676

Franschhoek Satellite Clinic
Akademie Street Franschhoek
021 876 2172

Groendal Kliniek
1 Stiebeuelstr, Franschhoek
021 876 3714

Idas Valley Clinic
Hammandshand Way Idas Valley, Stellenbosch
021 887 2721

Jamestown Clinic
Pajero Street, Jamestown, Stellenbosch
021 880 0357

Kayamandi Clinic
56 Bassi Street, Kayamandi, Stellenbosch
021 889 5061

Kylemore Clinic
19 School Street, Kylemore
021 885 2504

Rhodes Fruit Farm Clinic
Groot Drakenstein, 7680
021 874 1112

Somerset West Clinic
28 Church Street, Somerset West
021 444 4452

Victoria Street Clinic
5 Victoria Street, Stellenbosch
021 887 2650

People with weak immune systems - including **babies and people living with HIV** - have the highest risk of complications associated with measles infection. Vaccination is strongly recommended for these individuals. Pregnant women should not get vaccinated.

SACEMA
DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis

NRF National Research Foundation

UNIVERSITEIT STELLENBOSCH UNIVERSITY

science & technology
Department:
Science and Technology
REPUBLIC OF SOUTH AFRICA

Masels uitbraak in Stellenbosch

Vir meer inligting, stuur 'n e-pos na **measlesinfo@sacema.org** of skakel 021 808 2589.

'n SACEMA Inligtingsblad (Uitgereik 08 Feb 2017)

'n Aantal masels gevalle is bevestig in Stellenbosch, en die gevalle neem toe. **Wat moet jy weet?**

Inenting beskerm jou teen masels. Indien jy en jou kind ingeënt is, kan julle steeds masels kry, maar met minder ernstige simptome.

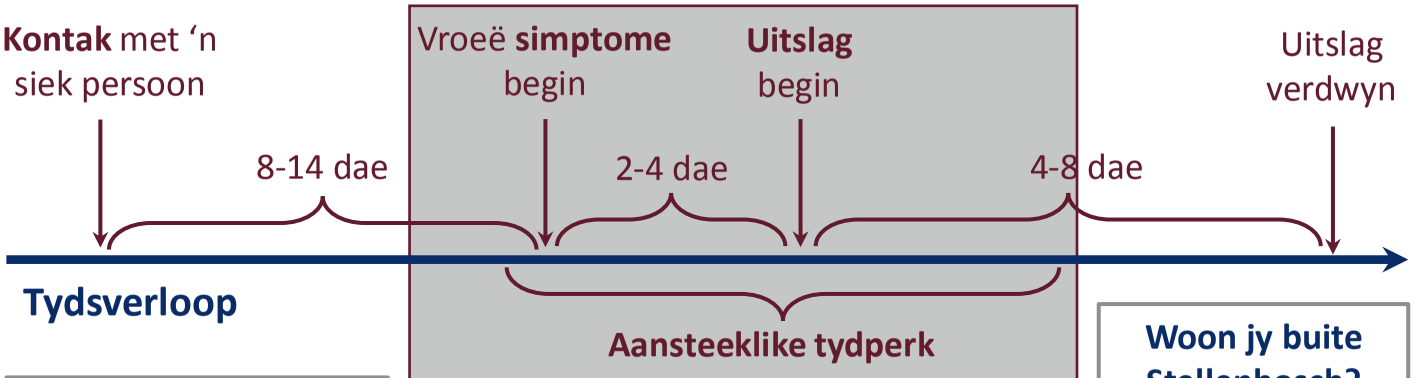
Wat kan ek doen?

As jy of iemand wat jy ken masels simptome het, bel jou dokter of plaaslike kliniek. Stel die gesondheidspersoneel vooraf in kennis dat jy masels vermoed. Vermy kontak met ander mense tot vier dae na die uitslag verskyn het. Hou siek kinders by die huis tot vier dae na die uitslag verskyn het.

Indien jy nie simptome het nie, word inenting aanbeveel – veral as jy nooit voorheen die inenting gekry het nie, as jy in kontak was met iemand met masels, of as jou immuunstelsel verswak is. Swanger vroue kan nie ingeënt word nie.

Plaaslike openbare klinieke verskaf gratis inentings vir kinders en grootmense.

Inentings kos R100-R220 by plaaslike apteke.



Hoe versprei dit?

Masels word versprei deur **kontak**. Dit sluit in aanraking, inasem van virusdeeltjies in die lug, of die aanraking van besmette oppervlaktes. Simptome sal eers 8-14 dae na kontak met 'n siek persoon verskyn.

Wanneer kan iemand met masels iemand anders aansteek? Een tot twee dae voor vroeë simptome begin en tot 4 dae na die uitslag verskyn.

Woon jy buite Stellenbosch?

Die uitbraak kan jou steeds affekteer. Masels is een van die mees aansteeklike siektes en versprei vinnig!

Wat is die simptome?

Vroeë **simptome** is soortgelyk aan verkoue en sluit in hoes, nies en rooi oë, asook hoë koors. Na 2-4 dae, sal 'n uitslag verskyn. Die **uitslag** sal van die kop en nek na die hele lyf versprei.

SACEMA DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis
NRF National Research Foundation
UNIVERSITEIT STELLENBOSCH UNIVERSITY
science & technology Department: Science and Technology REPUBLIC OF SOUTH AFRICA

Masels uitbraak in Stellenbosch

'n SACEMA Inligtingsblad (Uitgereik 08 Feb 2017)

Vir meer inligting, stuur 'n e-pos na **measlesinfo@sacema.org** of skakel 021 808 2589.

Plaaslike openbare klinieke verskaf gratis inentings vir kinders en grootmense.
Hier is 'n lys van klinieke in en rondom Stellenbosch.

Cloetesville Gemeenskaps Kliniek
Tennantville Straat, Stellenbosch
021 883 2676

Franschhoek Satelliet Kliniek
Akademie Straat, Franschhoek
021 876 2172

Groendal Kliniek
1 Stiebeuel Straat, Franschhoek
021 876 3714

Idas Vallei Kliniek
Hammandshand Way, Stellenbosch
021 887 2721

Jamestown Kliniek
Pajero Straat, Stellenbosch
021 880 0357

Kayamandi Kliniek
56 Bassi Straat, Stellenbosch
021 889 5061

Kylemore Kliniek
19 Skool Straat, Kylemore
021 885 2504

Rhodes Fruit Farm Kliniek
Groot Drakenstein
021 874 1112

Somerset Wes Kliniek
28 Kerk Straat, Somerset West
021 444 4452

Victoria Straat Kliniek
5 Victoria Straat, Stellenbosch
021 887 2650

Mense met verswakte immuunstelsels - insluitend **mense met MIV en babas** – het verhoogde risiko om komplikasies a.g.v. masels te kry. Inenting word sterk vir hierdie mense aanbeveel. **Swanger vroue** kan nie ingeënt word nie.

SACEMA
DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis

NRF National Research Foundation

UNIVERSITEIT STELLENBOSCH UNIVERSITY

science & technology
Department:
Science and Technology
REPUBLIC OF SOUTH AFRICA

Uqhambuko lwemasisi e-Stellenbosch

Ukuze ufumane ingcaciso ethe vetshe thumela i-imeyili ku: **measlesinfo@sacema.org** okanye tsalela umnxeba kwinombolo engu-021 808 2589.

Uxwebhu lwengcaciso lwe-SACEMA (Lwapapashwa ngowesi-08 kuFebruwari 2017)

Ugonyo luyawanciphisa amathuba okuhlaselwa kwakho yimasisi. Nokuba ugonyiwe, usenako ukugula, kodwa iimpawu zokugula aziyikuxhomisa amehlo kangango.

Ndingenza ntoni?

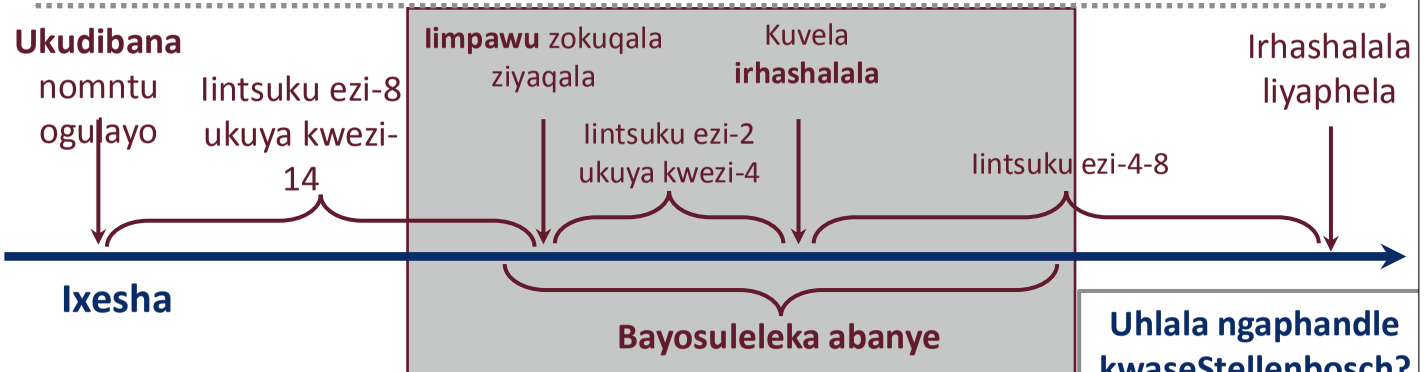
Ukuba wena okanye omnye umntu omaziyo uneempawu zemasisi, qhagamshelana nogqirha wakho okanye ikliniki ekufutshane. Ngaphambi kokuya kwagqirha okanye ekliniki, yazisa abasebenzi bezempilo ukuba kungenzeka ukuthi kanti unemasisi Musa ukusondela kwabanye abantu de kudlule iintsuku ezine emva kokuvela kwerhashalala. Abantwana abagulayo bagcine ekhaya bangayi esikolweni de kudlue iintsuku ezine emva kokuvela kwerhashalala.

Ukuba akubonkalisi zimpawu zokugula ucetyiswa ukuba ugonywe – ngakumbi xa wawungazange ugonywe, xa wawukhe wakufuphi nomntu onemasisi okanye ukuba isakhono somzimba wakho sokulwa nezifo senziwe sabuthathaka. Amabhinqa akhulelweyo makangagonywa.

Iikliniki **zoomasipala** zigonya iintsana simahla kambe zizinike nestofu sokuzomeleza simahla**.** Zona iikhemisti zalapha zibonelela ngogonyo lwemasisi nge-R100 ukuya kuma-R220.

Kubhaqwe imasisi eStellenbosch, kwaye inani labanayo liyenyuka.

Yintoni omawuyazi?



Yanda njani?

Imasisi idluliselwa ngoku**dibana,** nto leyo inokubandakanya ukubambana, ukuphefumlelana, okanye ukubamba iindawo ezi ngcolisekileyo. Iimpawu zibonakala kwiintsuku ezisi-8 ukuya kwezili-14 emva kokudibana nomntu ophethwe yiyo.

Umntu angabosulela nini abanye ngemasisi? Kusuku olunye ukuya kwezimbini ngaphambi kokuvela kweempawu zokuqala ukuya kuthi xhaxhe kwiintsuku ezine emva kokuvela kwerhashalala.

Uhlala ngaphandle kwaseStellenbosch?

Usengachaphazeleka. Imasisi sesinye sezona zifo zosulelayo kwaye sinaba ngokukhawuleza!

What are the symptoms?

Iimpawu zokuqala ziyafana nezentloko ebangwa ngumkhuhlane kwaye zingabandakanya ukukhohlela, ukuthimla namehlo abomvu, nto ezo ziphelekwa yifiva emandla. Emva kweentsuku ezi-2 ukuya kwezi-4 kuye kuvele irhashalala. **Irhashalala** elo linaba ukusuka apha entloko nasentanyeni ukuya emzimbeni wonke.

SACEMA DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis

National Research Foundation

UNIVERSITEIT STELLENBOSCH UNIVERSITY

science & technology
Department: Science and Technology
REPUBLIC OF SOUTH AFRICA

Uqhambuko lweMasisi e-Stellenbosch

Uxwebhu lwengcaciso lwe-SACEMA (Lwapapashwa ngowesi-08 kuFebruwari 2017)

Ingcaciso ethe vetshe thumela i-imeyili ku: **measlesinfo@sacema.org** okanye tsalela ku: 021 808 2589.

Iikliniki zoomasipala zigonya iintsana simahla kambe zizinike nestofu sokuzomeleza simahla

Khangela ikliniki ekufutshane kuwe

Cloetesville Community Health Clinic
Tennantville Street, Stellenbosch
021 883 2676

Franschhoek Satellite Clinic
Akademie Street Franschhoek
021 876 2172

Groendal Kliniek
1 Stiebeuelstr, Franschhoek
021 876 3714

Idas Valley Clinic
Hammandshand Way Idas Valley, Stellenbosch
021 887 2721

Jamestown Clinic
Pajero Street, Jamestown, Stellenbosch
021 880 0357

Kayamandi Clinic
56 Bassi Street, Kayamandi, Stellenbosch
021 889 5061

Kylemore Clinic
19 School Street, Kylemore
021 885 2504

Rhodes Fruit Farm Clinic
Groot Drakenstein, 7680
021 874 1112

Somerset West Clinic
28 Church Street, Somerset West
021 444 4452

Victoria Street Clinic
5 Victoria Street, Stellenbosch
021 887 2650

Abantu abamizimba isakhono sokulwa nezifo sibuthathaka – kubandakanywa **iintsna nabantu abaphila ne-HIV** – ngabona basemngciphekweni wokuba neengxaki ezinokoyanyaniswa nokosulelwa yimasisi. Abo bantu bacetyiswa kakhulu ukuba bagonywe. Amabhinqa amaithiyo makangagonywa..

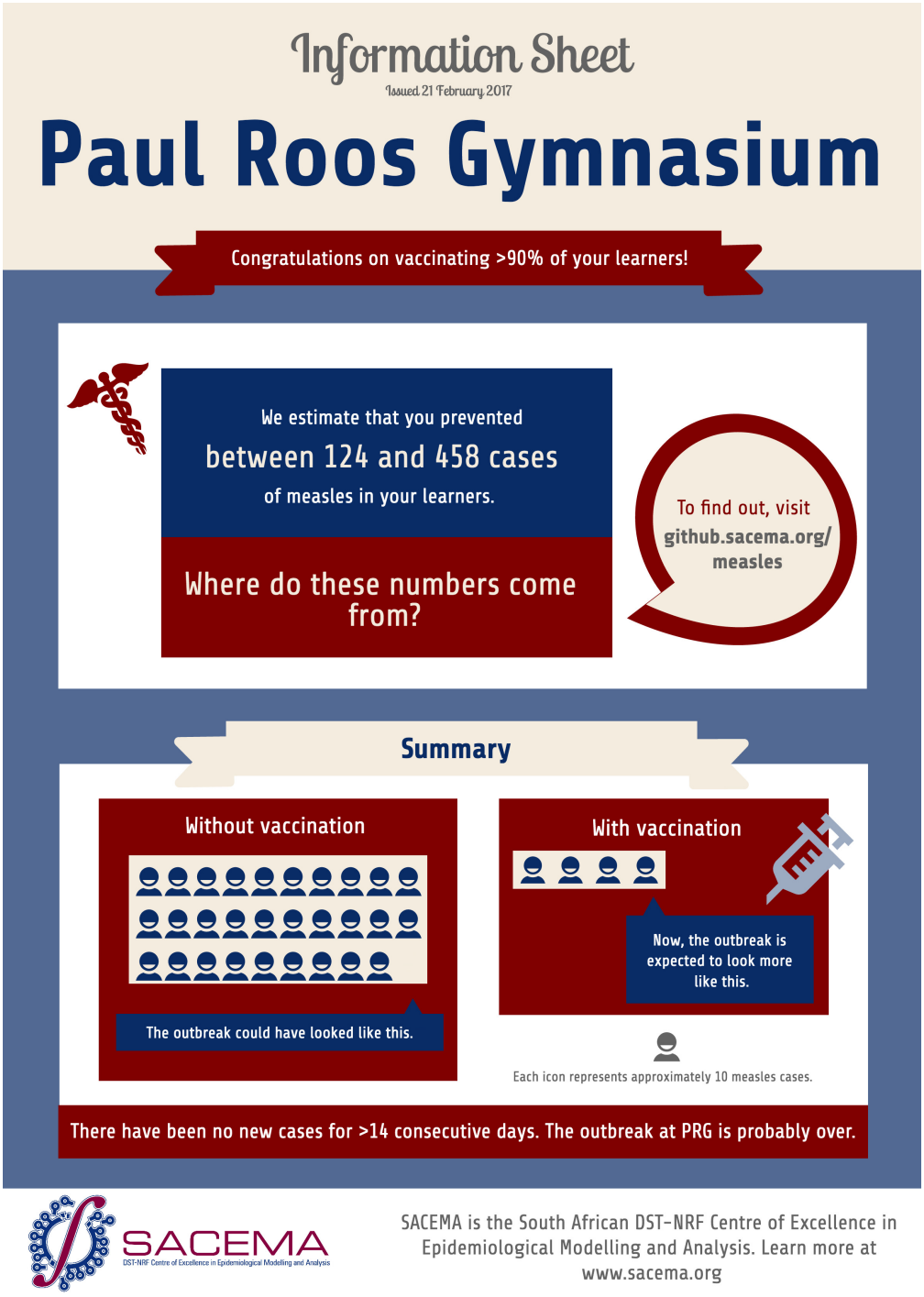
SACEMA
DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis

NRF National Research Foundation

UNIVERSITEIT STELLENBOSCH UNIVERSITY

science & technology
Department:
Science and Technology
REPUBLIC OF SOUTH AFRICA

A.2 Epidemic update sheet in English



List of references

Aaby, P., Knudsen, K., Jensen, T.G., Thirup, J., Poulsen, A., Sodemann, M., da Silva, M.C. & Whittle, H. 1990. Measles incidence, vaccine efficacy, and mortality in two urban African areas with high vaccination coverage. *Journal of infectious diseases*, 162(5):1043–1048.

Anderson, R.M., May, R.M. & Anderson, B. 1992. *Infectious diseases of humans: dynamics and control*, volume 28. Wiley Online Library.

Aylward, R.B., Clements, J. & Olive, J.-M. 1997. The impact of immunization control activities on measles outbreaks in middle and low income countries. *International journal of epidemiology*, 26(3):662–669.

Bailey, N.T. 1975. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.

Becker, A.D. & Grenfell, B.T. 2017. tsir: An R package for time-series susceptible-infected-recovered models of epidemics. *PloS one*, 12(9):e0185528.

Bellan, S.E., Pulliam, J.R., Pearson, C.A., Champredon, D., Fox, S.J., Skrip, L., Galvani, A.P., Gambhir, M., Lopman, B.A., Porco, T.C. *et al.*. 2015. Statistical power and validity of ebola vaccine trials in Sierra Leone: a simulation study of trial design and analysis. *The Lancet Infectious Diseases*, 15(6):703–710.

Biellik, R., Madema, S., Taole, A., Kutsulukuta, A., Allies, E., Eggers, R., Ngcobo, N., Nxumalo, M., Shearley, A., Mabuzane, E. *et al.*. 2002. First 5 years of measles elimination in Southern Africa: 1996–2000. *The Lancet*, 359(9317):1564–1568.

Bjørnstad, O.N., Finkenstädt, B.F. & Grenfell, B.T. 2002. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series sir model. *Ecological Monographs*, 72(2):169–184.

Black, R.E., Cousens, S., Johnson, H.L., Lawn, J.E., Rudan, I., Bassani, D.G., Jha, P., Campbell, H., Walker, C.F., Cibulskis, R. *et al.*. 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *The lancet*, 375(9730):1969–1987.

Blake, F.G. & Trask Jr, J.D. 1921. Studies on measles: III. acquired immunity following experimental measles. *The Journal of experimental medicine*, 33(5):621.

Bobashev, G.V., Ellner, S.P., Nychka, D.W. & Grenfell, B.T. 2000. Reconstructing susceptible and recruitment dynamics from measles epidemic data. *Mathematical Population Studies*, 8(1):1–29.

Bolker, B. & Grenfell, B. 1993. Chaos and biological complexity in measles dynamics. *Proceedings of the Royal Society of London B: Biological Sciences*, 251(1330):75–81.

Cairns, K.L., Perry, R.T., Ryman, T.K., Nandy, R.K. & Grais, R.F. 2011. Should outbreak response immunization be recommended for measles outbreaks in middle-and low-income countries? an update. *The Journal of infectious diseases*, 204(suppl_1):S35–S46.

Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. 2017. *Shiny: Web Application Framework for R*. R package version 1.0.5.
Available at: https://CRAN.R-project.org/package=shiny

Chiew, M., Gidding, H.F., Dey, A., Wood, J., Martin, N., Davis, S. & McIntyre, P. 2014. Estimating the measles effective reproduction number in Australia from routine notification data. *Bulletin of the World Health Organization*, 92(3):171–177.

Cori, A., Ferguson, N.M., Fraser, C. & Cauchemez, S. 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology*, 178(9):1505–1512.

Dalziel, B.D., Bjørnstad, O.N., van Panhuis, W.G., Burke, D.S., Metcalf, C.J.E. & Grenfell, B.T. 2016. Persistent chaos of measles epidemics in the prevaccination United States caused by a small change in seasonal transmission patterns. *PLoS computational biology*, 12(2):e1004655.

De Swart, R. 2009. Measles studies in the macaque model. *Curr Top Microbiol Immunol*, 330:55–72.

Dorélien, A.M., Ballesteros, S. & Grenfell, B.T. 2013. Impact of birth seasonality on dynamics of acute immunizing infections in sub-Saharan Africa. *PLoS One*, 8(10):e75806.

Earn, D.J., Rohani, P., Bolker, B.M. & Grenfell, B.T. 2000. A simple model for complex dynamical transitions in epidemics. *Science*, 287(5453):667–670.

Enders, J.F., Katz, S.L. & Holloway, A. 1962. Development of attenuated measles virus vaccines: A summary of recent investigation. *American Journal of Diseases of Children*, 103(3):335–340.

Ferrari, M.J., Bjørnstad, O.N. & Dobson, A.P. 2005. Estimation and inference of R0 of an infectious pathogen by a removal method. *Mathematical biosciences*, 198(1):14–26.

Ferrari, M.J., Fermon, F., Nackers, F., Llosa, A., Magone, C. & Grais, R.F. 2014. Time is (still) of the essence: quantifying the impact of emergency meningitis vaccination response in Katsina State, Nigeria. *International health*, 6(4):282–290.

Ferrari, M.J., Grais, R.F., Bharti, N., Conlan, A.J., Bjørnstad, O.N., Wolfson, L.J., Guerin, P.J., Djibo, A. & Grenfell, B.T. 2008. The dynamics of measles in sub-Saharan Africa. *Nature*, 451(7179):679–684.

Fine, P.E. 2003. The interval between successive cases of an infectious disease. *American journal of epidemiology*, 158(11):1039–1047.

Fine, P.E. & Clarkson, J.A. 1982. Measles in England and Wales–I: an analysis of factors underlying seasonal patterns. *International journal of Epidemiology*, 11(1):5–14.

Finkenstädt, B.F. & Grenfell, B.T. 2000. Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(2):187–205.

Fulginiti, V. 1965. Simultaneous measles exposure and immunization. *Archives of Virology*, 16(1):300–304.

Glass, K. & Grenfell, B. 2003. Antibody dynamics in childhood diseases: waning and boosting of immunity and the impact of vaccination. *Journal of theoretical biology*, 221(1):121–131.

Grais, R., De Radiguès, X., Dubray, C., Fermon, F. & Guerin, P. 2006*a*. Exploring the time to intervene with a reactive mass vaccination campaign in measles epidemics. *Epidemiology & Infection*, 134(4):845–849.

Grais, R.F., Conlan, J.K., Ferrari, M.J., Djibo, A., Le Menach, A., Bjørnstad, O.N. & Grenfell, B.T. 2007. Time is of the essence: exploring a measles outbreak response vaccination in Niamey, Niger. *Journal of the Royal Society, Interface / the Royal Society*, 5(18):67–74. ISSN 1742-5689.

Grais, R.F., Ferrari, M., Dubray, C., Bjørnstad, O., Grenfell, B., Djibo, A., Fermon, F. & Guerin, P.J. 2006*b*. Estimating transmission intensity for a measles epidemic in Niamey, Niger: lessons for intervention. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(9):867–873.

Grenfell, B.T., Bjørnstad, O.N. & Finkenstädt, B.F. 2002. Dynamics of measles epidemics: scaling noise, determinism, and predictability with the tsir model. *Ecological Monographs*, 72(2):185–202.

Griffin, D. & Pan, C.-H. 2009. Measles: old vaccines, new vaccines. In *Measles*, pages 191–212. Springer.

Griffin, D.E. & Oldstone, M.B. 2008. *Measles: Pathogenesis and control*, volume 330. Springer Science & Business Media.

Guerra, F.M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S.L., Li, Y. & Crowcroft, N.S. 2017. The basic reproduction number (R0) of measles: a systematic review. *The Lancet Infectious Diseases*. ISSN 1473-3099.
Available at: http://dx.doi.org/10.1016/S1473-3099(17)30307-9

Halloran, M.E., Auranen, K., Baird, S., Basta, N.E., Bellan, S., Brookmeyer, R., Cooper, B., DeGruttola, V., Hughes, J., Lessler, J. *et al.*. 2017. Simulations for designing and interpreting intervention trials in infectious diseases. *bioRxiv*, page 198051.

Heffernan, J. & Keeling, M.J. 2008. An in-host model of acute infection: Measles as a case study. *Theoretical population biology*, 73(1):134–147.

Heymann, D.L. 2004. *Control of communicable diseases manual 18th Edition*. American Public Health Association.

Hobson, F. 1934. The use of convaslescent measles serum in school epidemics. *The Lancet*, 224(5808):1408–1411.

Hobson, F. 1938. Measles: The conduct of a school epidemic. *The British Medical Journal*, pages 171–175.

Keeling, M.J. & Grenfell, B. 1997. Disease extinction and community size: modeling the persistence of measles. *Science*, 275(5296):65–67.

Keeling, M.J. & Rohani, P. 2008. *Modeling infectious diseases in humans and animals*. Princeton University Press.

Klepac, P., Bjørnstad, O.N., Metcalf, C.J.E. & Grenfell, B.T. 2012. Optimizing reactive responses to outbreaks of immunizing infections: balancing case management and vaccination. *PloS one*, 7(8):e41428.

Krugman, S., Giles, J.P., Friedman, H. & Stone, S. 1965. Studies on immunity to measles. *The Journal of pediatrics*, 66(3):471–488.

Lessler, J., Metcalf, C.J.E., Cutts, F.T. & Grenfell, B.T. 2016. Impact on epidemic measles of vaccination campaigns triggered by disease outbreaks or serosurveys: a modeling study. *PLoS medicine*, 13(10):e1002144.

Markowitz, L.E., Preblud, S.R., Fine, P.E. & Orenstein, W.A. 1990. Duration of live measles vaccine-induced immunity. *The Pediatric infectious disease journal*, 9(2):101–110.

Metcalf, C., Lessler, J., Klepac, P., Cutts, F. & Grenfell, B. 2012. Impact of birth rate, seasonality and transmission rate on minimum levels of coverage needed for rubella vaccination. *Epidemiology & Infection*, 140(12):2290–2301.

Metcalf, C.J.E., Bjørnstad, O.N., Grenfell, B.T. & Andreasen, V. 2009. Seasonality and comparative dynamics of six childhood infections in pre-vaccination Copenhagen. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1676):4111–4118.

Minetti, A., Hurtado, N., Grais, R.F. & Ferrari, M. 2013. Reaching hard-to-reach individuals: nonselective versus targeted outbreak response vaccination for measles. *American journal of epidemiology*, 179(2):245–251.

Moss, W.J. & Griffin, D.E. 2012. Measles review article, Moss & Griffin. *The Lancet*, 379:153–164.

Mossong, J. & Muller, C.P. 2003. Modelling measles re-emergence as a result of waning of immunity in vaccinated populations. *Vaccine*, 21(31):4597–4603.

Mossong, J., Nokes, D.J., Edmunds, W.J., Cox, M.J., Ratnam, S. & Muller, C.P. 1999. Modeling the impact of subclinical measles transmission in vaccinated populations with waning immunity. *American Journal of Epidemiology*, 150(11):1238–1249.

Orenstein, W.A. & Gay, N.J. 2004. The theory of measles elimination: implications for the design of elimination strategies. *The Journal of infectious diseases*, 189(Supplement_1):S27–S35.

Orenstein, W.A., Perry, R.T. & Halsey, N.A. 2004. The clinical significance of measles: a review. *The Journal of infectious diseases*, 189(Supplement_1):S4–S16.

Perry, R.T., Gacic-Dobo, M., Dabbagh, A., Mulders, M.N., Strebel, P.M., Okwo-Bele, J.-M., Rota, P.A., Goodson, J.L. *et al.*. 2014. Progress toward regional measles elimination worldwide, 2000–2013. *MMWR Morb Mortal Wkly Rep*, 63(45):1034–8.

Pinner, R.W., Onyango, F., Perkins, B.A., Mirza, N.B., Ngacha, D.M., Reeves, M., DeWitt, W., Njeru, E., Agata, N.N., Broome, C.V. *et al.*. 1992. Epidemic meningococcal disease in Nairobi, Kenya, 1989. *Journal of infectious diseases*, 166(2):359–364.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
Available at: https://www.R-project.org/

Shibeshi, M.E., Masresha, B.G., Smit, S.B., Biellik, R.J., Nicholson, J.L., Muitherero, C., Shivute, N., Walker, O., Reggis, K. & Goodson, J.L. 2014. Measles resurgence in Southern Africa: challenges to measles elimination. *Vaccine*, 32(16):1798–1807.

Uzicanin, A. & Zimmerman, L. 2011. Field effectiveness of live attenuated measles-containing vaccines: a review of published literature. *The Journal of infectious diseases*, 204(suppl_1):S133–S149.

Wallinga, J., Levy-Bruhl, D., Gay, N. & Wachmann, C. 2001. Estimation of measles reproduction ratios and prospects for elimination of measles by vaccination in some Western European countries. *Epidemiology & Infection*, 127(2):281–295.

WHO. 1999. WHO guidelines for epidemic preparedness and response to measles outbreaks.

WHO. 2005. Givs: global immunization vision and strategy: 2006-2015.

WHO. 2009. Response to measles outbreaks in measles mortality reduction settings.

WHO. 2017. Measles vaccines: WHO position paper (april 2017). *Weekly Epidemiological Record*, 92(17):205–227.
Available at: http://apps.who.int/iris/bitstream/10665/255377/1/WER9217-205-227.pdf

Yeung, L.F., Lurie, P., Dayan, G., Eduardo, E., Britz, P.H., Redd, S.B., Papania, M.J. & Seward, J.F. 2005. A limited measles outbreak in a highly vaccinated US boarding school. *Pediatrics*, 116(6):1287–1291.